REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	Decmber 1997	End of Year

4. TITLE AND SUBTITLE
A Hierarchical Neural Network Based Data Processing System for Ground-Penetrating Radar, An End of Year report for Project CU/1049/6: Application of Neural Networks Coupled with Genetic Algorithms to Optimize Soil Cleanup Operations in Cold Climates

5. FUNDING NUMBERS
N/A

6. AUTHOR(S)
John M. Sullivan

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Remote Sensing/GIS Division
US Army Cold Regions Research and
Engineering Laboratory
Hanover, NH 03755

8. PERFORMING ORGANIZATION REPORT NUMBER
N/A

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
SERDP
901 North Stuart St. Suite 303
Arlington, VA 22203

10. SPONSORING / MONITORING AGENCY REPORT NUMBER
N/A

11. SUPPLEMENTARY NOTES
This work was funded in part by SERDP Project 1049/6 and by CRREL SFRC Number DACA89-97-K0001.

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release: distribution is unlimited.

12b. DISTRIBUTION CODE
A

13. ABSTRACT ***(Maximum 200 Words)***

Ground-Penetrating Radar (GPR) is a powerful modern tool to examine the structure and properties of the media below the ground surface within a depth of 30 meters. This study is very important for the environmental problems related to the transport of contaminants in ground waters. Successful GPR profiling of the subsurface media yielding the correct information about the distribution of permafrost, water table, and bedrock depths is the key factor in ground water flow modeling.

This work attempts to develop a hierarchical processing system capable of handling GPR data characterized by high degree of uncertainty, natural physical ambiguity, and, sometimes, missing or incorrect entries. The hierarchical nature of the algorithm allows to split the task of media profiling into several consecutive stages, each of the following has less degree of uncertainty than the previous one. Neural Networks modules are designed to accomplish the two main processing goals of recognizing the "subsurface pattern" followed by the identification of the depths of the subsurface layers like permafrost, groundwater table, and bedrock. Pre-processing procedure has the objective to transform raw GPR data into a small feature vecor containing the most representative and discriminative features of the signal. The feature vector coupled with other relevant GPR information forms the input for the Neural Network processing units.

14. SUBJECT TERMS
ground-penetrating radar
GPR
neural networks
remote sensing
SERDP
SERDP Collection

15. NUMBER OF PAGES
90

16. PRICE CODE N/A

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
unclass.	unclass.	unclass.	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

19990521 144

SERDP Report

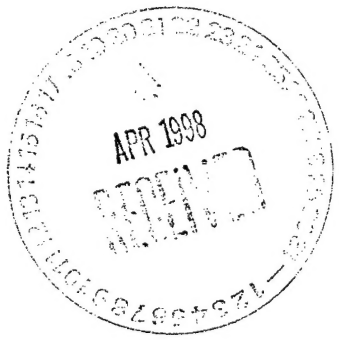
A Hierarchical Neural Network Based Data Processing System for Ground-Penetrating Radar

An End of Year report for Project CU/1049/6:

Application of Neural Networks Coupled with Genetic Algorithms to Optimize Soil Cleanup Operations in Cold Climates

Submitted to:

Dr. Femi Ayorinde
SERDP Program Manager for Cleanup
901 North Stuart Street, Suite 303
Arlington, VA 22203



by

Dr. John M. Sullivan, Jr.

Remote Sensing/GIS Division
US Army Cold Regions Research and Engineering Laboratory
Hanover, NH 03755

December 1997

ABSTRACT

Ground-Penetrating Radar (GPR) is a powerful modern tool to examine the structure and properties of the media below the ground surface within a depth of 30 meters. This study is very important for the environmental problems related to the transport of contaminants in ground waters. Successful GPR profiling of the subsurface media yielding the correct information about the distribution of permafrost, water table, and bedrock depths is the key factor in ground water flow modeling.

This work attempts to develop a hierarchical processing system capable of handling GPR data characterized by high degree of uncertainty, natural physical ambiguity, and, sometimes, missing or incorrect entries. The hierarchical nature of the algorithm allows to split the task of media profiling into several consecutive stages, each of the following has less degree of uncertainty than the previous one.

Neural Networks modules are designed to accomplish the two main processing goals of recognizing the "subsurface pattern" followed by the identification of the depths of the subsurface layers like permafrost, ground water table, and bedrock. Pre-processing procedure has the objective to transform raw GPR data into a small feature vector containing the most representative and discriminative features of the signal. The feature vector coupled with other relevant GPR information forms the input for the Neural Network processing units.

The separate system components are implemented in software and tested with artificial as well as real GPR data. The entire processing system is trained with synthetic inputs and tested with real measured data. The algorithm demonstrates correct, accurate, and tolerant to noise/error performance. This establishes the feasibility of the system application to true-to-life problems. The developed software proved to be fast enough to consider the possibility of real-time implementation for field operation in conjunction with the industrial GPR equipment.

ACKNOWLEDGMENTS

This work was funded in part by SERDP (Strategic Environmental Research and Development Program) Project 1049/6 cleanup thrust area and by CRREL (U.S. Army Cold Regions Research and Engineering Laboratory) SFRC Number DACA89-97-K0001.

This December 1997 document integrates all work performed to date on this project (CU/1049/6). This work has been presented as a Master Thesis (with credit and acknowledgements to SERDP) by D. Repin under the Academic direction of the project PI, Dr. John M. Sullivan, Jr.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1 GPR technique

Ground Penetrating Radar (GPR) is an electromagnetic remote sensing technique which uses radio waves, typically in the 10 to 2500 MHz frequency range, to locate and map different features and structures below the ground surface (bgs). In general, a GPR system transmits a short electromagnetic pulse into the ground - the pulse is reflected, refracted or scattered by the targets that exhibit some difference in electrical properties (dielectric permittivity, conductivity, and magnetic permeability) and is then recorded by the receiving antennas. The greater is the difference in the dielectric permeability, the larger is the amplitude of the reflection pulse.

High radar frequencies are needed to achieve a good spatial resolution, but penetration depth of the electric field is inversely proportional to the frequency. Hence the choice of frequency range is a trade-off between resolution and penetration depth. Penetration depth also depends on the nature of the soil, which has different attenuation properties. For example, desert sand has an attenuation of about 1 dB/m for a 1 GHz frequency, clay has an attenuation of 100 dB/m at the same frequency.

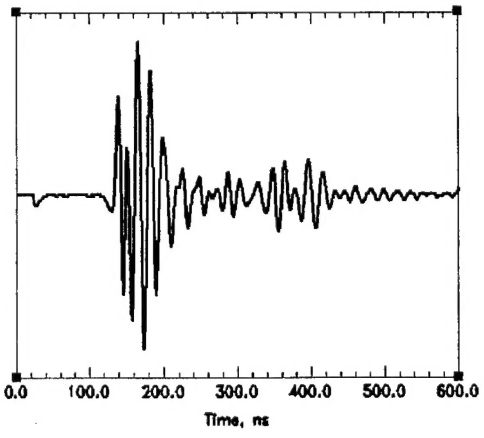


Figure 1-1. Typical GPR A-scan.

The reflected wave is sampled and digitized by an A/D converter to form a vector. Typically 512 or 1024 points are taken through the region of interest. The recorded signal in the time domain is called an A-scan and it may look similar to that shown in the Figure 1-1. In many GPR applications the A-scans are recorded consecutively along some spatial direction usually called radar or transect line.

Typical GPR system records 5 to 10 scans per meter. A variety of software packages exist for visualization and data processing (see Section 1.3 for more details).

Typical GPR soundings are performed by dragging a GPR hardware package including transmitting and receiving antennas behind a vehicle - Figure 1-2.

Figure 1-2. Field measurements with the GPR hardware dragged behind a vehicle.

1.2 Application of the GPR

GPR technique includes various scientific, industrial, environmental and military applications, among those are:

- Stratigraphic layers profiling (water table detection, etc.);
- Ice thickness measurements;
- Buried objects detection;
- Archaeological investigations;
- Road investigations;
- Fracture detection in hard rock;
- Liquid contamination detection;
- Safety inspections at nuclear power plants;
- Anti-personnel mines detection.

This work concentrates on a particular application - non-invasive site characterization of stratigraphic layer depths in the vicinity of Fairbanks, Alaska, USA. One of the most important features of this region is the presence of the permanently frozen materials (permafrost), which have distinct dielectric properties. The GPR data were collected to identify horizontal and vertical distributions of permafrost zones, water table and bedrock. GPR was chosen to accomplish this task since [1]: the scales of depths and lateral variations of permafrost are too small to resolve with seismic methods, and too large for efficient mapping with electromagnetic inductance methods. Frozen soils (mostly sands and gravels) are a low loss propagation media for electromagnetic waves at radio frequencies. Therefore, saturated sediments form a continuous and highly reflective surface with frozen ground. In this situation GPR is one of the best tools to study the vertical distribution of the different subsurface materials.

The ability to characterize the subsurface media is very important for the environmental problems related to transporting of contaminants with ground water. Successful GPR profiling plays a critical role in the site description.

1.3 Previous research efforts

Significant research efforts in the field of GPR were made by CRREL (US Army Cold Regions Research and Engineering Laboratory). They performed Multi-bandwidth reflection profiling of discontinuous permafrost via GPR during 1993-94. The GPR antennas bandwidths centered near 50, 100 and 300 MHz. An area spanning 8 km^2 had over 100 km of GPR profiles recorded [1].

The subsurface layers configurations for the area is shown schematically in Figure 1-3. Variations in the permafrost structure may have natural (presence of the river) or artificial (roads or other human activities) origins, and those variations may be significant over a relatively short lateral scale. On the other hand variations of the water table and the bedrock absolute depth are relatively small. Along with the above mentioned major subsurface components variations in types of soils and moisture content introduce additional difficulties into appropriate data processing.

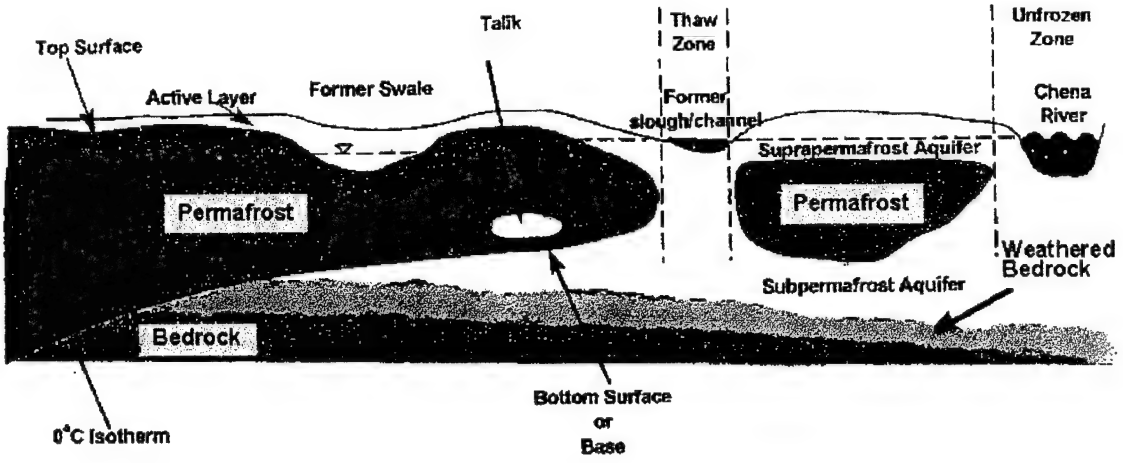


Figure 1-3. Possible configuration of subsurface layers [1].

The commercial software package RADAN™ by Geophysical Survey Systems, Inc. was used to view the collected GPR data. The reflection time for one layer of an electromagnetic pulse from the interface of materials with the different dielectric properties can be expressed as:

$$t_{refl} = \frac{2\sqrt{\varepsilon}d}{c} \tag{1.1}$$

where c is the light speed, d is the layer thickness, and ε is the corresponding dielectric constant.

Previously GPR line data processing used RADAN™ software to visualize sets of adjacent A-scans coupled with a human expert evaluation based on previous knowledge and generalization from the available geophysical data (Figure 1-4). This analysis was augmented with the information of the subsurface structure from a number of boreholes. Those boreholes were drilled throughout the area, and types and properties of materials as well as corresponding layers depths were recorded.

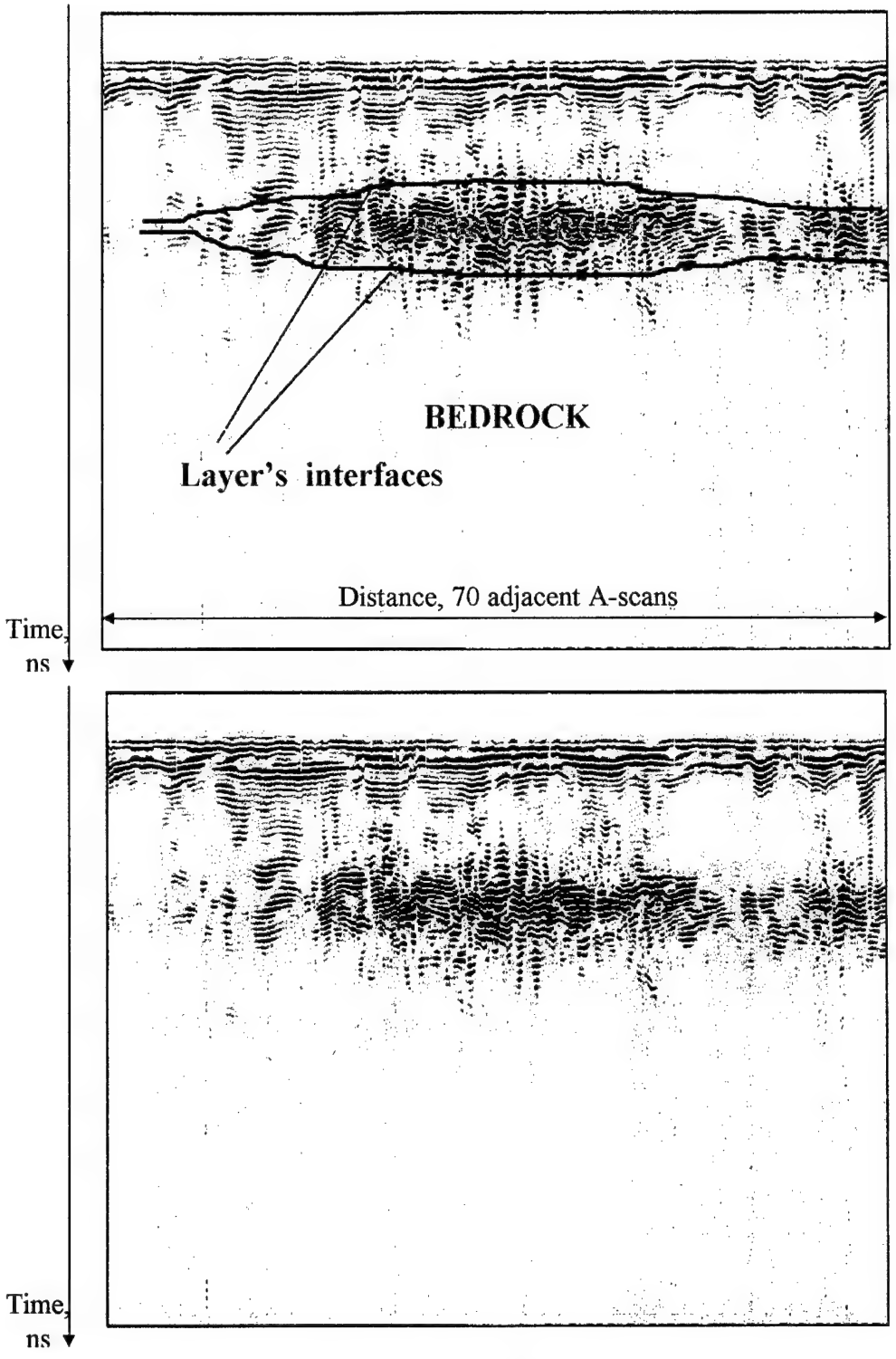


Figure 1-4. A schematic manual identification of subsurface features (top), the same part of the GPR line without any marks (bottom).

The previously mentioned method is highly subjective. It lacks any form of automation and the layer interfaces identified by the expert still lack depth information. The time scale of the GPR signal does not scale linearly with depth. That is why a new approach has been developed for automatic stratigraphic layer identification and depths prediction.

1.4 Hierarchical neural network based approach

The proposed processing architecture has the following advantageous features:

- main task is split into several consecutive stages decreasing the degree of uncertainty within each step;
- data adaptive techniques are applied to the GPR signal transforming it into a highly informative and easily interpretable feature vector;
- neural network modules used provide high error and noise tolerance.

A block diagram of the approach is shown in Figure 1-5. The major processing units of the data processing system are neural networks (NN). One NN (NN1) is designed to perform best on a classification problem. The second NN (NN2) provides good approximation ability for depth analysis within the classified problem. Both neural networks belong to the class of two-layer feedforward networks. Both neural networks are of the Multi-Layer Perceptron (MLP) type trained with Backpropagation or Scaled Conjugate Gradient algorithms.

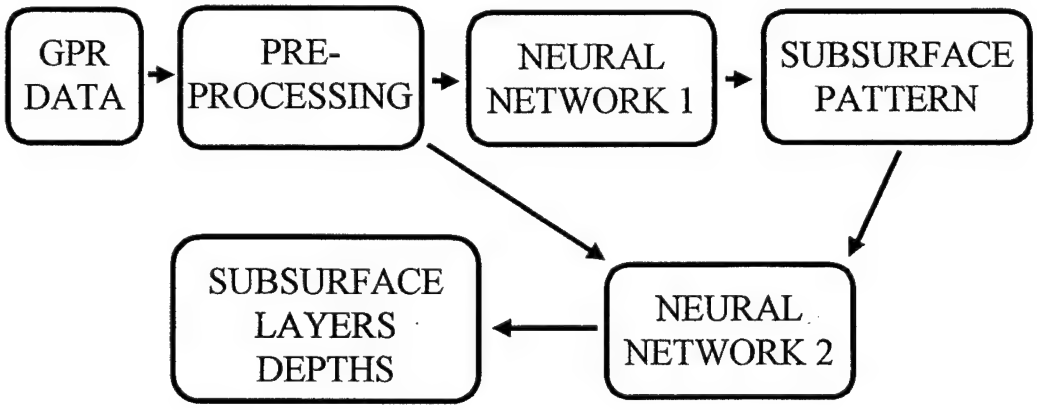


Figure 1-5. Block diagram of the data processing architecture.

Another important part of this architecture is the pre-processing stage - it has four separate blocks responsible for proper data handling and feature extraction to provide the

neural networks with relevant input information.

In general the system operation may be described as: the pre-processing unit performs checking of the data consistency; the second stage decomposes the initial A-scan with our Adaptive Transform (AT) technique into a set of data adaptive basis functions; the third section incorporates information about the previous A-scans with the linear regression method into the current processing step. Then a feature vector formed from the coefficients of the AT decomposition and some other prior available information is fed into the Neural Network 1 (NN1) to recognize the "subsurface pattern" for the current A-scan. This subsurface pattern reflects one of the possible subsurface layers configuration (similar to Figure 4-3). During the next step part of the transformed input for the NN1 and already available information about the subsurface pattern is used as input for Neural Network 2 (NN2), which produces the set of the startigraphic layers depths.

A feedback routine is used to account for possible incorrect recognition of the pattern or other "alarm" signals produced during system operation. This feedback feature adds flexibility to the entire set-up that already has as high noise/error tolerant processing units as neural networks.

1.5 Outline

The introduction briefly described the basics of the GPR technique, prospectives of its application to the stratigraphic layers profiling, some of the previous research efforts, and the general approach, which is used in this work to address the problem under consideration.

Chapter 2 of this work gives an overview of the available experimental (real) GPR and supplementary (boreholes) information along with the brief description of the Finite Difference Time Domain numerical simulations for the Plane Wave formulation. These data are used in all stages of the system development: pre-processing, training and results verification.

Chapter 3 provides a theoretical background for the Multilayer Perceptron neural network models The chapter covers basic principles, variations of training algorithms, operating modes and details of the implementation.

Chapter 4 discusses the pre-processing stage. Implications for choosing the particular pre-processing algorithms like dimension reduction and discriminative feature extraction are reflected. It also considers several possible alternatives (Fourier Transform, Wavelet Transform) to the Adaptive Transform and the logical chain of the concept improvement.

Chapter 5 presents details of the hierarchical GPR data processing systems, which incorporates methods and techniques described in Chapters 3 and 4. The issues of parameters choice and optimization, accuracy and reliability of the entire system are discussed. Results of the tests for different kinds of data are evaluated.

Chapters 6 summarizes the current stage results and describes possible further work. Appendix A provides additional information about simulated GPR lines and tests performed. Appendix B gives a detailed classification of the subsurface patterns and their description as they were used for simulations along with examples of the typical A-scans.

2. REAL AND SIMULATED GPR DATA

2.1 Soil properties and layers nomenclature

The term "Ground Penetrating Radar" implies the use of radar signals directed below the ground surface. This section summarizes factual as well as some unsubstantiated knowledge about the electrical properties for several types of soils and soil structures. The properties that characterize the propagation of the electromagnetic waves in an isotropic medium are the dielectric permittivity {ε} and the electric conductivity {σ}. The first one is responsible for the wave propagation time, the second one - for the loss factor.

According to various references [1-3] both values are very sensitive to the soil moisture content. For example, ε for dry soil can be in a range of 4 - 8, but with a moisture content of 30% ε can be as high as 40. The soil conductivity affects mostly the EM pulse evanescence and has a small effect on the reflections off the layers boundaries. Discussion of its properties is postponed to Section 2.3.2. Dielectric permittivity values for the most types of soils (dry, saturated, frozen, unfrozen) are presented in the Table 2.1.

Table 2.1. Soil dielectric permittivity.

SOIL TYPE	PERMTTIVITY VALUE
Permanently frozen materials	4.4 - 5.6
Unfrozen saturated sediments	12.0 - 45.0
Dry soil	4.0 - 8.0
Weathered bedrock	> 11.0
Bedrock: granite, sandstone	7.0 - 9.0, 10.0
Sand: dry, 15% moisture, 25% moisture content	≈3.0, ≈9.0, ≈25.0

In GPR applications the following classifications are used (see Figure 1-4), where "soils" in this work refer to sands and gravels, which constitute much of the subsurface media in Fairbanks, AK:

- Active Layer - first layer of soil under the surface with some organics present;
- Water Table (WT) - the top boundary of the ground water;
- Permafrost - permanently frozen soils;
- Suprapermafrost Aquifer - saturated soils above permafrost;
- Subpermafrost Aquifer - saturated soils below permafrost;
- Talik - unfrozen zone within permafrost;
- Thaw - zone, where permafrost is absent mostly due to the artificial reasons;
- Weathered Bedrock - zone, where bedrock is not solid and is mixed with soil;
- Bedrock - the continental platform.

2.2 *Experimental GPR and related data*

2.2.1 GPR hardware system

Ground-Penetrating Radar by Geophysical Survey Systems, Inc. SIR Model 4800 was primarily used during the field measurements. Available collection rate ranged from 10 to 50 scans/second. Two antennas were used for transmitting and receiving the signal. The transmitter antenna radiated a broadband wavelet pulse with a duration ranging from about 1.5 to 2.5 cycles. An identical, but separate, antenna was used as a receiver, because echoes could return from near surface events before the transmitter antenna had stopped radiating. A fiber optic connection was used to trigger the transmission. Measurements with wavelet bandwidths centered near 50, 100, and 300 MHz were used employing a custom made antenna and industrial antennas GSSI Model 3207, GSSI Model 3102, respectively. Antennas were dragged approximately 10 meters behind a vehicle moving at a speed about 1 meter/second.

2.2.2 A-scans, radar lines, borehole data

A-scans or simply scans are the natural output of the GPR system. A set of a few hundred or thousands adjacent A-scans collected with a frequency of 5-10 scans per meter is called a radar line (GPR line). Data were stored in binary files, which contained some additional information about the GPR system configuration and parameters at the

time of the measurements - the header (see Figure 2-1). A fixed number of 512 (sometimes 1024) sampling points were used in the field experiments. Duration of the scan could vary from 500 to 1000 ns but remained the same for a single radar line. Amplitude of the reflections was sampled as an 8 or 16-bit signal.

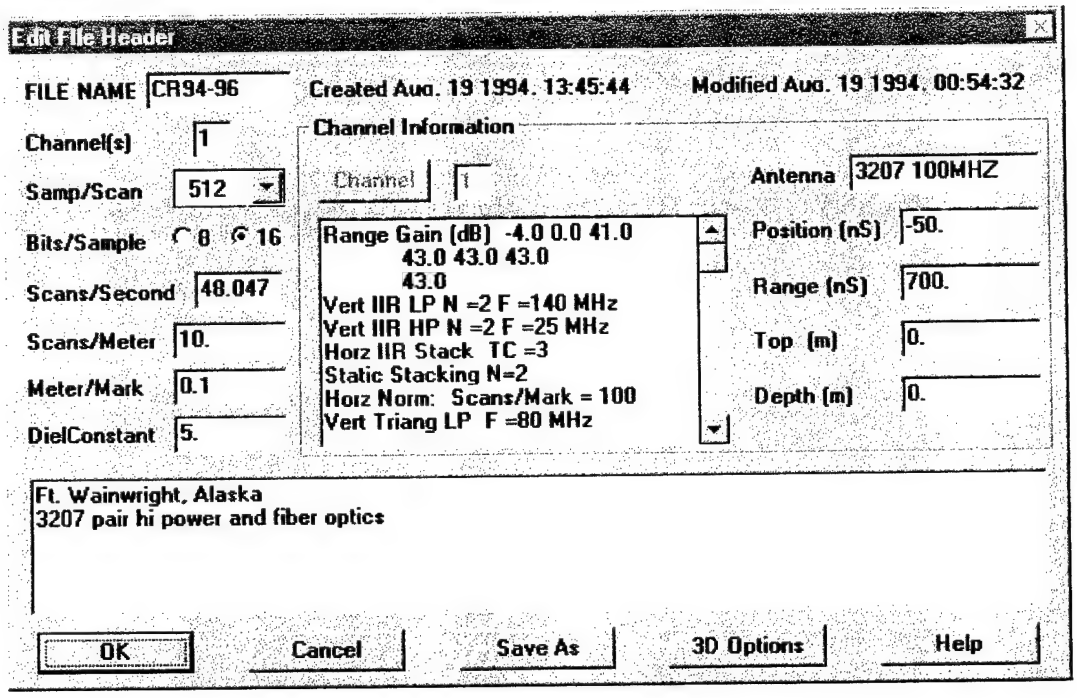


Figure 2-1. GPR line file header content.

Information about more than 50 boreholes which were made for different purposes (research, monitoring) was available. Most of those boreholes had corresponding GPR information. Depths of the boreholes varied from 60 to 170 ft, drilling was usually terminated at bedrock or up to 45 ft below permafrost in unfrozen sediments.

Figure 2.2 shows boreholes corresponding to various GPR lines. The Groundwater Modeling Software 2.0 (GMS) was used to process the area map with the GPR lines (white image in the upper part) and the finite element mesh of the same area. After transforming the image in a way that the river contour of the map coincides with the river elements in the mesh the state plane coordinates become available to locate the beginning and the end of the radar line (x, y). Then a simple routine automatically scans

through the entire set of the boreholes checking their coordinates and picking those that are close to the radar line of interest.

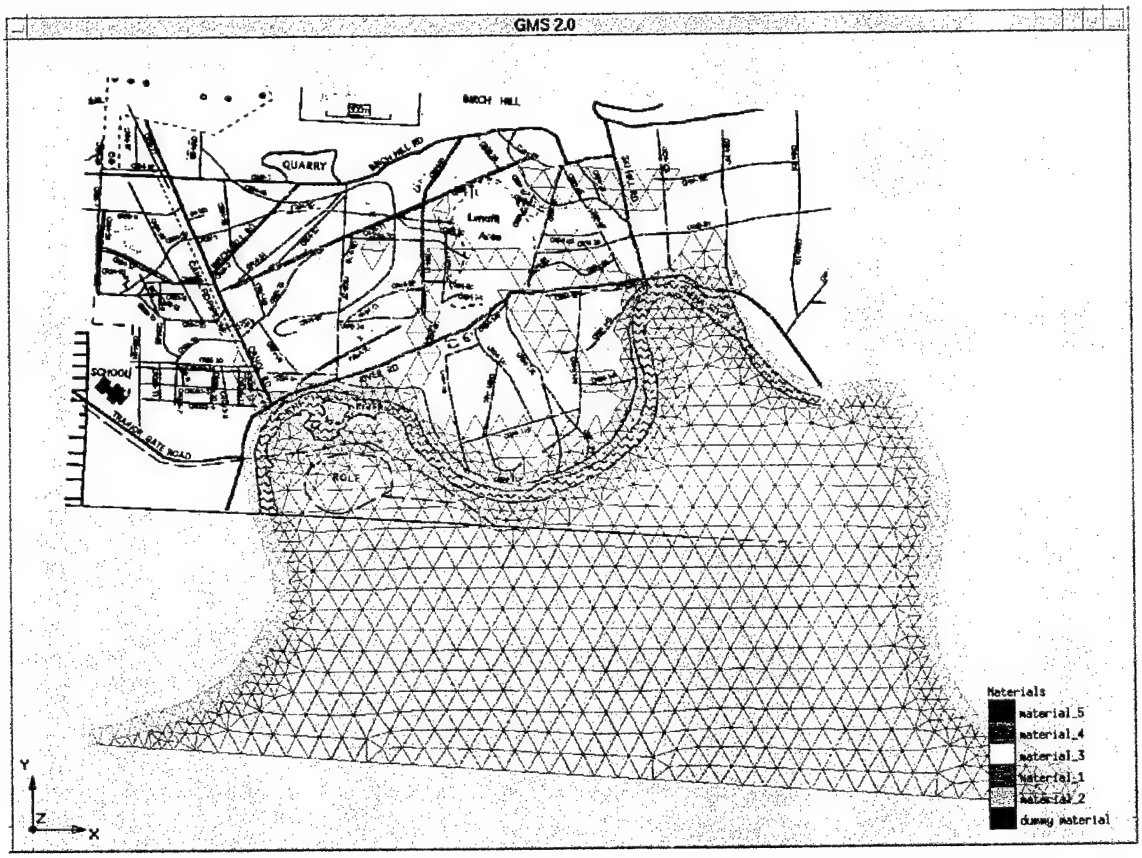


Figure 2-2. Draft area map and the GMS mesh superposition.

Two figures below show the examples of the A-scans from radar line CR94-61r. The space locations of those scans correspond to the boreholes so that the conclusions about the nature of the reflection pulses can be made.

The scan in Figure 2-3 (total time - 1000 ns) has two clear groups of reflections, according to the available borehole data (Tables 2.2 - 2.4) the first one is likely to be associated with the top of permafrost at the depth of 4 ft bgs, another reflection is probably due to the subpermafrost aquifer or intrapermafrost inclusions.

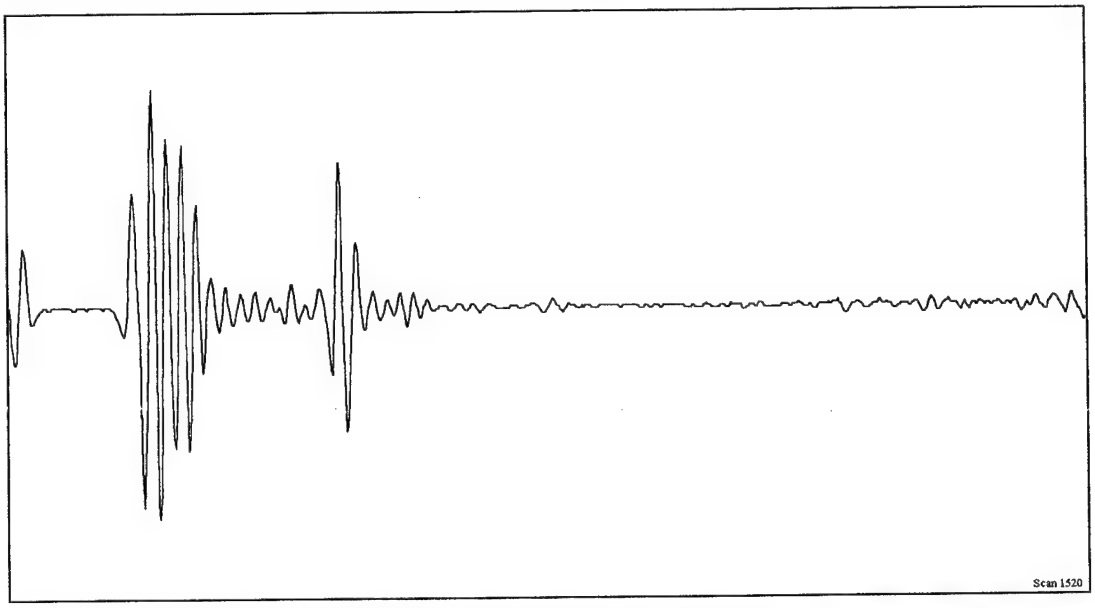


Figure 2-3. A-scan 1520 (1000 ns, 1024 sampling points).

Table 2.2. Boring information (including WT), scan 1520.

Well Identifier	X state plane	Y state plane	Surface elevation, ft	Depth of the hole, ft	Weather cond.	Depth to WT	Date
AP-6551	246401.385	3969487.000	443.50	38.0	Pt. Cld,45F	N/A	9/15/94

Table 2.3. Lithology information, scan 1520.

Well Identifier	Layer index (from top of hole)	depth to layer top (ft)	depth to layer bottom (ft)	Primary Lithology	Lithology modifier
AP-6551	005	20.00	38.00	SAND	SILTY
AP-6551	004	15.50	20.00	SAND	POORLY GRADED
AP-6551	003	7.50	15.50	SAND	SILTY
AP-6551	002	0.50	7.50	SILT	
AP-6551	001	0.00	0.50	PEAT	

Table 2.4. Permafrost information, scan 1520.

Well Identifier	Layer number (in sequence)	Depth below ground surface to top of layer (feet)	Depth below ground surface to bottom of layer (feet)	Zone Identifier
AP-6551	001	0.00	4.00	UNFROZEN
AP-6551	002	4.00	38.00	FROZEN

The A-scan in Figure 2-4 has overlapping reflections due to the presence of several zones with and without permafrost and the ground water table: transitions from frozen to unfrozen and especially to saturated zone cause strong reflection patterns. Borehole information related to this scan location is presented in Tables 2.5-2.7.

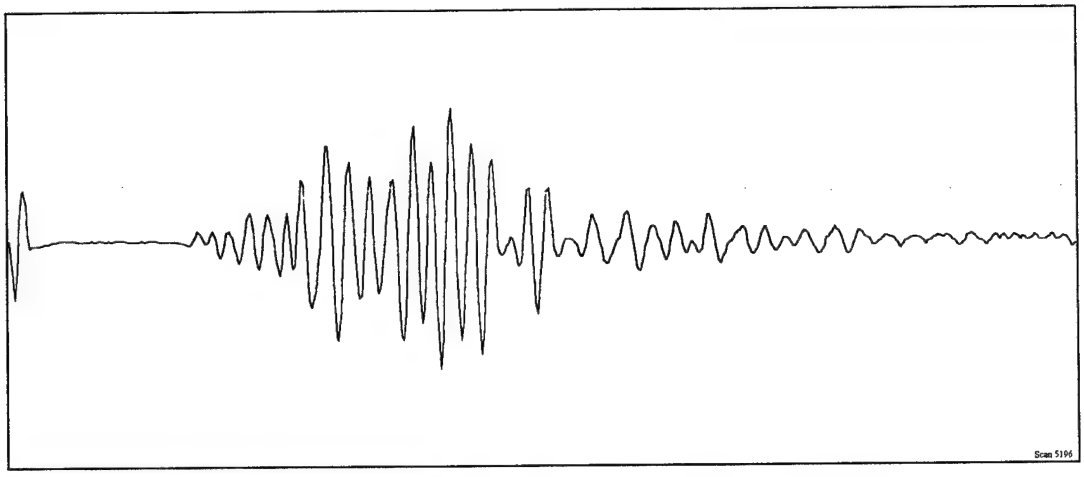


Figure 2-4. A-scan 5196 (600 ns, 512 sampling points).

Table 2.5. Boring information (including WT), scan 5196.

Well Identifier	X state plane	Y state plane	Surface elevation, ft	Depth of the hole, ft	Weather cond.	Depth to WT	Date
AP-6223	244727.672	3969556.25	441.30	93.00	Cloudy,25 F	8.0	11/8/93

Table 2.6. Lithology information, scan 5196.

Well Identifier	Layer index (from top of hole)	depth to layer top (ft)	depth to layer bottom (ft)	Primary Lithology	Lithology modifier
AP-6223	010	77.50	93.00	SAND	POORLY GRADED
AP-6223	009	60.00	77.50	SAND	SILTY
AP-6223	008	40.50	60.00	GRAVEL	POORLY GRADED
AP-6223	007	32.50	40.50	GRAVEL	SILTY
AP-6223	006	25.50	32.50	GRAVEL	POORLY GRADED
AP-6223	005	21.00	25.50	SAND	POORLY GRADED
AP-6223	004	16.00	21.00	GRAVEL	POORLY GRADED
AP-6223	003	11.00	16.00	SAND	POORLY GRADED
AP-6223	002	3.00	11.00	GRAVEL	POORLY GRADED
AP-6223	001	0.00	3.00	SILT	

Table 2.7. Permafrost information, scan 5196.

Well Identifier	Layer number (in sequence)	Depth below ground surface to top of layer (feet)	Depth below ground surface to bottom of layer (feet)	Zone Identifier
AP-6223	001	0.00	1.00	FROZEN
AP-6223	002	1.00	38.50	UNFROZEN
AP-6223	003	38.50	86.00	FROZEN
AP-6223	004	86.00	93.00	UNFROZEN

2.2.3 Major difficulties with data processing

A fundamental problem for the GPR data processing lies in the reflection time versus ε-depth relation (Equation 1.1): for a fixed reflection time there exists an almost infinite (limited only by the physical constraints) number of the combinations of layer depth and corresponding dielectric permittivity values. Our processing system strives to resolve this ambiguity, but fails at times due to poorly collected data, improperly set gain control or other unknown conditions (see Figure 2-5).

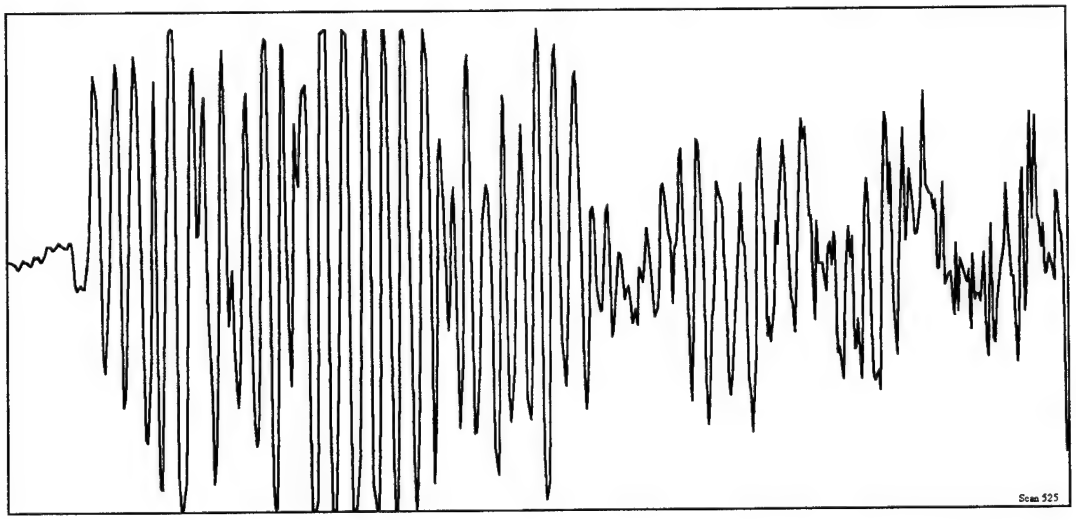


Figure 2-5. Uncertain A-scan due to improper gain settings (amplitude "chopping" example).

Frequently, an 8-bit representation of the signal was not sufficient for certain cases. Some of these problems could be resolved: for example, with some prior

knowledge the "chopped" signal may be rescaled and using the carrier frequency value chopped spikes can be restored. But the concept of the automatic layers profiling system implies no human control of the signal before it goes into the processing system.

2.3 Simulated FDTD data

For better understanding of the underlying physics associated with the electromagnetic pulse propagation through the ground, a Finite Difference Time Domain (FDTD) simulator was developed [19]. Several physical models were examined: plane wave (see Figure 2.6), line source, and finite aperture source configurations. Satisfactory agreement of simulated and real data was achieved for the plane-wave formulation.

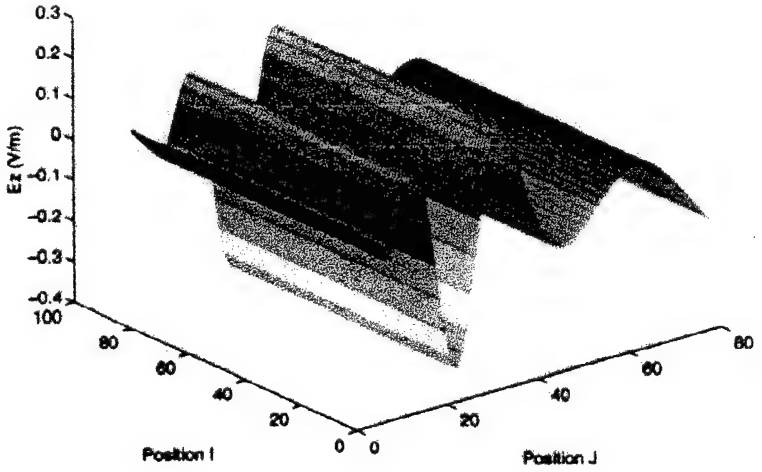


Figure 2-6. Simulated plane-wave electromagnetic pulse in 3D.

2.3.1 Scattering and loss

Assuming the plane-wave formulation represented the signal behavior in the subsurface medium, there was a concern to account for some physical concepts such as scattering and loss in the numerical model.

Scattering may be responsible for two possible ways of distorting the signal with respect to the uniform lossless medium model: a change of the dielectric constant due to the presence of the scattering particles and an energy dissipation due to the scattering

process [9]. Loss may be considered as an absorption effect in the medium incorporating the imaginary part of the dielectric constant.

A change of the dielectric constant may be expressed as: $\varepsilon_{effective} = \varepsilon \cdot \alpha(\varepsilon, n_0, r)$, where ε is dielectric constant without scattering, n_0 is the concentration of the spherical scattering particles of radius r. The change value for the case of $\varepsilon = 20.0$ is about 10% for sand/gravel soil. This effect was not implemented directly into the numerical model. Rather, the measured ε values already accounted for scattering and other factors [1] was used.

2.4 Matching experimental and simulated data

One of the chief reasons to use the simulated data, during the different stages of the entire processing system development, is the lack of the 100% reliable experimental GPR data. Even the presence of the borehole information leaves some degree of ambiguity while trying to assign certain ε values to the different subsurface layers. Simulated data allow one to figure out the origin of every single reflection: i.e. for the overlapping reflections it is much easier to verify the decomposition technique - amplitudes of the separate reflections are known in advance.

Another critical reason for choosing the synthetic data is the possibility of generating a training set for the neural network: this issue is discussed in details in Section 5.3. The ability to match real GPR scans gives the ground for the assumption that a correctly chosen simulated training set provides an accurate representation of the "entire space" of the possible GPR A-scans allowing to adequately process real GPR data.

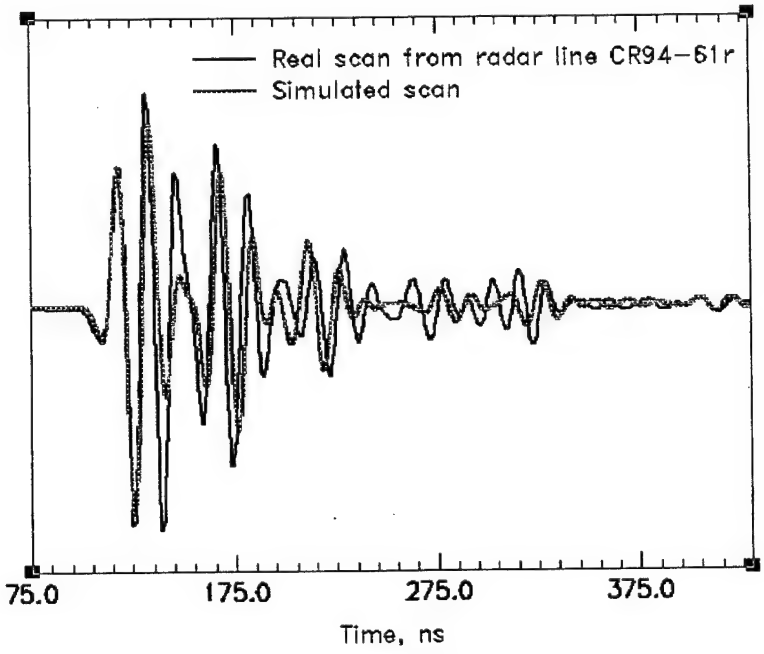


Figure 2-7. Real and simulated scans on the time scale.

Figure 2-7 shows a real scan from the CR94-61r line and the simulated counterpart. Relatively good agreement is observed taking into account some difference between the propagating EM pulse shape in the field measurements, and the synthetic initial pulse in FDTD model. The time-depth diagram for the same pair of scans is shown in Figure 2-8.

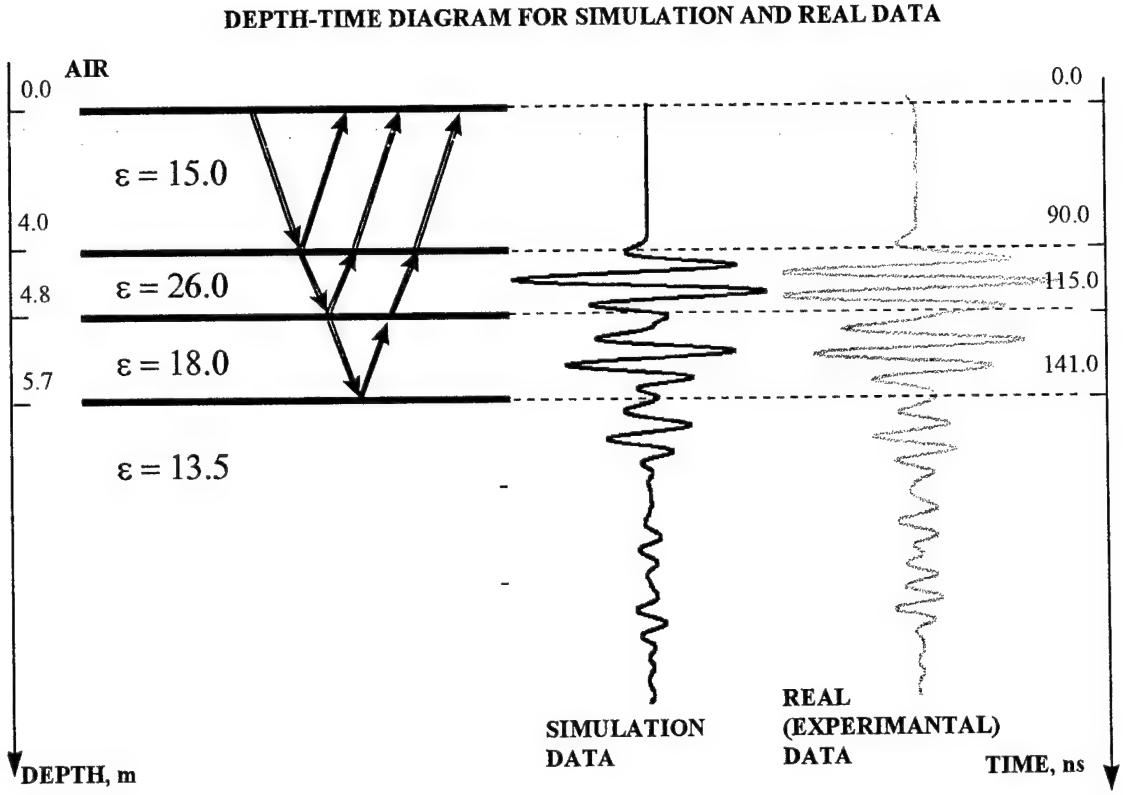


Figure 2-8. Depth-time diagram for real and simulated scans.

3. NEURAL NETWORKS FOR CLASSIFICATION AND APPROXIMATION PROBLEMS

Supervised neural networks find many applications in two general problem formulation: decision-based and approximation/optimization-based [6]. The decision-based problems as a final result aim at a certain organization of categories or classification of the patterns into some distinct groups. Categories are usually labeled as binary values and the training procedure is based on the correct classification for each training pattern, synaptic weights and biases of the network are adjusted to accomplish this task.

The approximation-based problem has the objective to map the input space onto the output using a finite set of data points and then to be able to get an output value for any datum from the input space as close to the correct value as possible. For this formulation outputs are usually analog and their exact values are important. Network parameters are modified during the training to minimize some energy function that depends on the difference of the actual and the desired network output.

Both formulations do not have significant differences in their algorithms and implementation, but certain details are changed to achieve the best performance for each task. In this work the first formulation is used for subsurface pattern identification, the second one for identification of the startigraphic layers depths.

3.1 MultiLayer Perceptron (MLP)

Multilayer perceptron is probably one the most well-known neural networks, that are trained with the supervised algorithms. It belongs to the class of feedforward layered neural networks - it has an input layer, output layer and one or several internal or hidden layers. Connections in this type of network exist only between the neurons of different layers and each neuron is connected to all the neurons in the previous and in the following layers. There are no lateral (within one layer) connections or feedback. The input signal propagates through the network in a forward direction on a layer-by-layer basis.

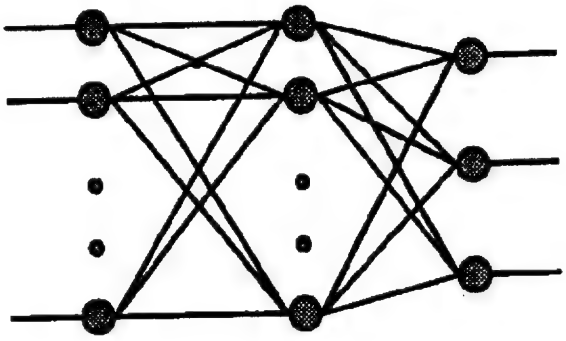


Figure 3-1. MLP network.

MLPs found their application in different technological and other areas. Each processing unit (neuron) of the MLP network can be described with an activation or transfer function, the essential feature of that function is non-linearity and for many applications it may similar to the Equations 3.1 or 3.2:

$$f(u_i) = \frac{c_1}{1+\exp\{-u_i\}}, \tag{3.1}$$

$$f(u_i) = c_1 \frac{u_i}{|u_i| + c_2}, \tag{3.2}$$

where c_1, c_2 are constants and u_i is the output of the neuron before applying the non-linearity:

$$u_i = \sum_{j=1}^{M} w_{ij} \cdot x_j + \theta_i, \tag{3.3}$$

where x_j is the network input or the output of the previous layer, w and θ are the weights and the bias corresponding to this neuron respectively. Input propagation through the network is a forward run, finally it produces a set of outputs - activation signals of the output layer units. During the forward run all the weight and biases values remain fixed. During the backward run weights and biases are adjusted according to the error-correction learning rule. The error is defined as a sum of squared differences of the activation values of output neurons and the target output values provided by the training set. The error propagates backward through the network and the weights and biases are adjusted to reduce the cumulative error. This algorithm is a "Backpropagation of error" learning algorithm.

In this work an MLP implementation with only one hidden layer is considered: there are two facts that proves the sufficiency of this assumption. The first is a theorem by A.N. Kolmogorov [14] that states that a two-layer (actually three input-hidden-output) network is able to perform mapping from an arbitrary input space onto the arbitrary output space, the only detail that is missing is the number of neurons in the hidden layer, but there exist a configuration which allow to perform a desired correct mapping. Another argument in support of the assumption above is that a considerable major of the current MLP applications are accomplished with only one hidden layer and acceptable results are reported in many cases.

3.2 Backpropagation learning algorithm

As far the derivation of the Backpropagation algorithm is presented in most of the neural network books [5], [6] only the brief summary is given here and the discussion is focused on the details of implementation, modifications of the algorithm and the modes of operation.

Let (x_i, y_i) be a single pattern in the training set, M is the number of training patterns. For a single neuron which output (u) and activation ($a=f(u)$) functions are defined as Equations 3.3 and 3.2 (3.2 is used instead of 3.1 to allow the activation to be in the range [-1.0,1.0]). For the input layer $a_i(0) \equiv x_i$, for the output layer $a_i(L) \equiv y_i$, L is the total number of layers, N_l- number of neurons in a layer.

The objective is to modify weights w_{ij} and biases θ_i to minimize the error energy function:

$$E = \frac{1}{2}\sum_{m}^{M}\sum_{i}^{N}[y_i^{(m)} - a_i^{(m)}(L)]^2 \tag{3.4}$$

During the backward run weight values (and biases also, but later on only the expression for weights are presented) are updated:

$$w_{ij}^{m+1}(l) = w_{ij}^{m}(l) + \Delta w_{ij}^{m}(l) \tag{3.5}$$

where the superscript indexes denote the discrete time step or the iteration and

$$\Delta w_{ij}^{m}(l) = -\eta \frac{\partial E}{\partial w_{ij}^{m}(l)} = \eta \delta_i^{(m)}(l) f'(u_i^{(m)}(l)) a_j^{(m)}(l-1), \tag{3.6}$$

$$\delta_i^{(m)}(l) = -\frac{\partial E}{\partial a_i^{(m)}(l)}, \tag{3.7}$$

where δ is the error signal, η is the learning rate term, and f' is the derivative of the activation function with respect to a single weight.

The keys to Backpropagation algorithm are the recursive formulas:

$$\delta_i^{(m)} = \sum \delta_i^{(m)}(l+1) f'(u_j^{(m)}(l+1)) w_{ij}^{(m)}(l+1) \tag{3.8}$$

$$w_{ij}^{(m+1)}(l) = w_{ij}^{(m)}(l) + \eta \delta_i^{(m)}(l) f'(u_i^{(m)}(l)) a_j^{(m)}(l-1) + \alpha \Delta w_{ij}^{(m+1)}(l) \tag{3.9}$$

Using this formulas error signal may be calculated recursively for any neuron and its weights and biases may be updated accordingly. The only modification for the standard Backpropagation is the *momentum term* - $\alpha \Delta w_{ij}^{(m+1)}(l)$. Using the momentum term implies adaptation of the learning step size with respect to the weight change on the previous iteration: so flat spots on the energy surface are traversed faster, and the step decreases for the rough spots.

The Backpropagation algorithm described above refers to the *on-line* mode when the weights are updated after presenting each single pattern. *Batch* mode has the difference that the weight are updated once in an epoch (a full presentation of all training patterns). For this mode all weights and biases changes are summed over one epoch.

In this work on-line mode of the Backpropagation algorithm was implemented (sometimes it is also called "vanilla backpropagation" [18]). But the supervised learning of the feedforward networks is not limited to the discussed methods. As far as the task is to minimize the error energy function, and the techniques above use only first order Gradient Descent method. So the task may be formulated as an unconstrained nonlinear function optimization problem [5]. The natural step forward is to use a second order or other modern procedures (like Conjugate Gradient) to accomplish the task of finding the

acceptable minimum on the error surface. This extension was made in the next section, where the Scaled Conjugate Gradient algorithm is discussed.

3.3 Scaled Conjugate Gradient method

Scaled Conjugate Gradient (SCG) algorithm for fast BP training was developed by M. Moller [7] in 1991 and is one of the most efficient in its class.

SCG belongs to the class of Conjugate Gradient methods and has the major advantage in the speed of convergence and the lack of the parameters to be defined by the user before the training procedure. Most of the algorithms require at least the initial learning rate value, momentum term, etc. SCG behaves in a similar way as the ordinary CG [4],[5], but has a built-in feature that allow to determine a step size (learning rate for the particular iteration) in an automatic manner at a relatively low computational cost. This technique is applicable to the *batch* mode of the BackPropagation algorithm.

The SCG is applied to the generalized weight vector consisting of all the weights and biases of the network (in this work we consider only a two-layer network):

$$\widetilde{w} = (.., w_{ij}^1, .., \theta_j^1, .., w_{jk}^2, .., \theta_k^2, ..),$$

where w denote the weights values, θ denote biases and superscript indexes refer to the layer number, subscript ones to the number of element in the layers, total number of components in this vector is N.

Network energy function is defined as the sum of error squares (Equation 3.4) and the derivative $E'(\widetilde{w}) = \left(.., \sum_{p=1}^{P} \frac{dE_p}{dw_{ij}^1}, .. \sum_{p=1}^{P} \frac{dE_p}{d\theta_j^1}, .. \sum_{p=1}^{P} \frac{dE_p}{dw_{jk}^2}, .. \sum_{p=1}^{P} \frac{dE_p}{d\theta_k^2}, .. \right)$,

where P is the total number of patterns in the training set and p refers to the E_p for the particular pattern. The main feature of the SCG is the approximation of the second order information:

$$\widetilde{s}_k = E''(\widetilde{w}_k) \cdot \widetilde{p}_k \approx \frac{E'(\widetilde{w}_k + \sigma_k \widetilde{p}_k) - E'(\widetilde{w}_k)}{\sigma_k},$$

where $\widetilde{p}_k$ is the conjugate direction, which is the key to compute the step size α_k for the each component of the weight vector updating - $\widetilde{w}_{k+1} = \widetilde{w}_k + \alpha_k \widetilde{p}_k$. Another critical

feature of this algorithm is making the Hessian matrix to be always positive definite by adjusting the scalar parameter λ_k in the expression below:

$$\tilde{s}_k = \frac{E'(\tilde{w}_k + \sigma_k \tilde{p}_k) - E'(\tilde{w}_k)}{\sigma_k} + \lambda_k \tilde{p}_k .$$

The SCG implementation may be summarized as the following:

1. Parameters initialization: $0 < \sigma < 10^{-4}$, $0 < \lambda_1 < 10^{-6}$, $\varepsilon_1 = 10^{-8}$, $\varepsilon_2 = 10^{-10}$ - these numbers are mostly determined by the precision of the machine, $\bar{\lambda}_1 = 0$, and the initial weight vector $\tilde{w}_1$ is chosen (random values are assigned to the weights and biases). Gradient direction $\tilde{p}_1 = \tilde{r} = -E'(\tilde{w}_1)$ is calculated, variable SUCCESS set to be equal "TRUE".
2. Second order information is calculated,

   $\sigma_k = \sigma / |\tilde{p}_k|$;

   $\tilde{s}_k = (E'(\tilde{w}_k + \sigma_k \tilde{p}_k) - E'(\tilde{w}_k)) / \sigma_k$;

   $\delta_k = \tilde{p}_k^T \tilde{s}_k$;
3. δ_k is scaled as $\delta_k = \delta_k + (\lambda_k - \bar{\lambda}_k)|\tilde{p}_k|^2$.
4. If $\delta_k \le 0$ then the Hessian matrix is made to be a positive definite:

   $\bar{\lambda}_k = 2(\lambda_k - \delta_k / |\tilde{p}_k|^2)$;

   $\delta_k = -\delta_k + \lambda_k |\tilde{p}_k|^2$;

   $\lambda_k = \bar{\lambda}_k$.
5. Step size is calculated:

   $\mu_k = \tilde{p}_k^T \tilde{r}_k$;

   $\alpha_k = \mu_k / \delta_k$.
6. Then the comparison parameter Δ is calculated:

   $\Delta_k = 2\delta_k \left[E(\tilde{w}_k) - E(\tilde{w}_k + \alpha_k \tilde{p}_k) \right] / \mu_k^2$.
7. If $\Delta_k \ge 0$, then the error can be reduced:

   $\tilde{w}_{k+1} = \tilde{w}_k + \alpha_k \tilde{p}_k$;

$\tilde{r}_{k+1} = -E'(\tilde{w}_{k+1})$;

$\bar{\lambda}_k = 0$, SUCCESS = TRUE;

If the total number of iterations so far has reached N - the algorithm is restarted:

$\tilde{p}_{k+1} = \tilde{r}_{k+1}$;

else:

$\beta_k = (|\tilde{r}_{k+1}|^2 - \tilde{r}_{k+1}^T \tilde{r}_k) / \mu_k$;

$\tilde{p}_{k+1} = \tilde{r}_{k+1} + \beta_k \tilde{p}_k$.

If $\Delta_k \geq 0.75$ then the scale parameter is reduced $\lambda_k = \frac{1}{4}\lambda_k$.

If $\Delta_k < 0$:

$\bar{\lambda}_k = \lambda_k$;

SUCCESS = FALSE.

8. If $\Delta_k < 0.25$ then the scale parameter is increased:

$\lambda_k = \lambda_k + (\delta_k(1-\Delta_k)/|\tilde{p}_k|^2)$.

9. If the steepest descent direction $\tilde{r}_k \neq \tilde{0}$, then $k = k+1$ and go to the step 2, otherwise the algorithm is terminated and the desired generalized weight vector $\tilde{w}_{k+1}$ is the final output. Another termination criterion which proved to be valuable to avoid floating-point exceptions is introduced in [8] and implemented in this work:

$2 \cdot |E(\tilde{w}_{k+1}) - E(\tilde{w}_k)| \leq \varepsilon_1(|E(\tilde{w}_{k+1})| + |E(\tilde{w}_k)| + \varepsilon_2)$

The SCG was not benchmarked against the regular Backpropagation in this work due to the lack of really large scale problems. For the details on comparison with the different techniques one may refer to [7].

4. PRE-PROCESSING FOR FEATURE EXTRACTION

For the many signal/data processing architectures a proper choice of the pre-processing technique is one of the most significant factors, which determines the performance of the entire system [4]. Several pre-processing procedures address the following goals.

- to eliminate inconsistent data from regular consideration;
- to eliminate the redundant information thus reducing the dimension of the problem;
- to enhance the most discriminative features of the different data categories;
- to scale or transform the data to allow the main processing system to operate in the most favorable regime.

From now on the discussion addresses the default GPR setup: a 600 ns A-scan recorded with a center frequency of 100 MHz, and the term "soil" refers mostly to sands and gravels, unless stated otherwise.

4.1 Correlation based pre-processing for checking data consistency

The goal of this technique is to determine inconsistent A-scans in order to exclude them from regular processing. Inconsistency means that the current A-scan does not fit the "trend" or appears to be out of the scan sequence. To accomplish this goal a cross-correlation operation is performed for every two adjacent scans and the space derivative of the correlation value is calculated. If $\left|d\Gamma/dr\right| >> 0$ - then an alarm signal is sent and the current scan is marked as suspect (Γ is the cross-correlation value, r is the distance along the radar line).

This operation is performed in the Fourier domain to achieve shift invariance. Small changes in the scan starting point, which could be the result of some bump on the road where the antenna is being dragged, may cause significant differences in Γ values

in the time domain, while the Fourier spectra of the adjacent scans remain almost the same.

Here $\Gamma = \sum_{i=1}^{N/2} F_i F_{i+1}$, where F is the magnitude of the corresponding Fourier coefficient, N is the number of sampling points in the scan. Figure 4-1 shows Γ and $\left| d\Gamma / dr \right|$ values for a sample A-scan sequence.

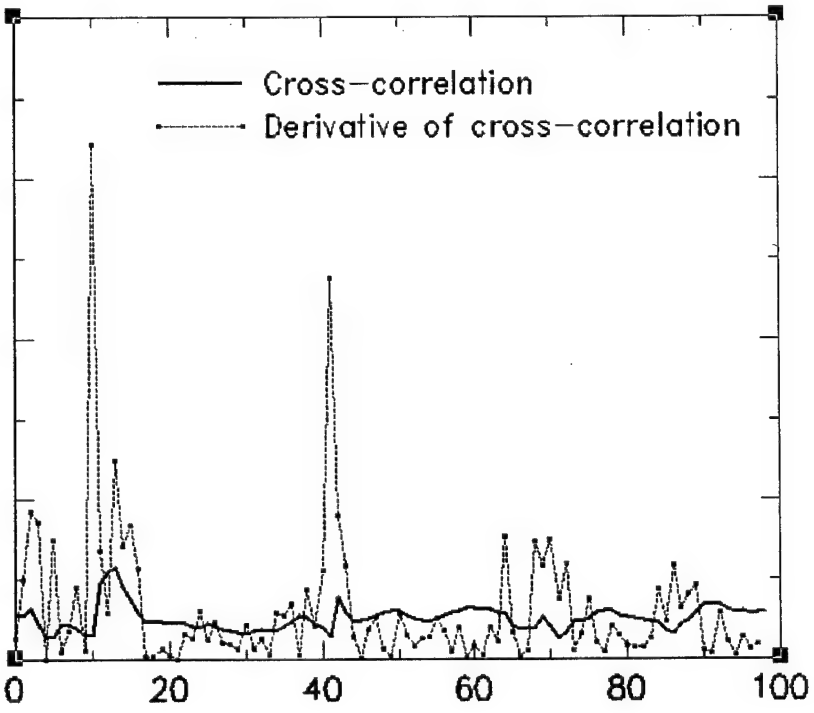


Figure 4-1. Cross-correlation and its derivative relative values vs. scan number.

4.2 Dimension reduction and discriminative features extraction

The main preprocessing stage of the data passed through the consistency checking routine addresses the following issues:

- reducing the dimension of the problem - a typical A-scan is represented with 512 - 1024 sampling points - too many for using as inputs for the neural;
- preserving the chronological order of reflection pulses contained in the A-scan;

- enhancing those features in the signal that reveal the most discriminative characteristics of the data which allow one to distinguish between different subsurface patterns;
- incorporating available prior knowledge about possible feature combinations imposing certain physical constraints (ε values, layering order).

The dimensional issue is very important in terms of the size of the data set for training the neural network: higher dimension produces larger neural network. When the input vector dimension reaches the value of N = 40 - 60, it becomes extremely difficult either to generate sufficient amount of synthetic data. The dependency may be expressed as follows: the required number of training patterns T should be much larger (5-10 times) than $(N{\cdot}M + M{\cdot}O)$ - the "total number of weights in the network" [20]. Here M is the number of neurons in the hidden layer, and O is the number of output neurons. For this particular problem M is often chosen to be about 10. That is why N of about 50 produces T on the order of 10^3 - 10^4 making the approach infeasible in terms of generation of the required number of training examples

4.3 Fourier Transform

Application of a Discrete Fourier Transform (DFT) to an A-scan takes advantage of the shift invariance property of the FT component magnitudes (discarding the phase) and thus representing the scan by a feature vector consisting of the magnitude values in the frequency domain alone. The general difficulty is that this procedure does not produce a unique mapping of the scan into the frequency domain, and two different scans may turn out to have very similar Fourier spectra.

Fourier spectra (256-point FT) of four real GPR A-scans are shown in Figure 4-2 (left) and one may see the apparent distinct features of each spectrum. However, feature vectors constructed from the FT magnitudes did not provide an accurate pattern recognition (the data were tested with the algorithm described in the Section 5.3).

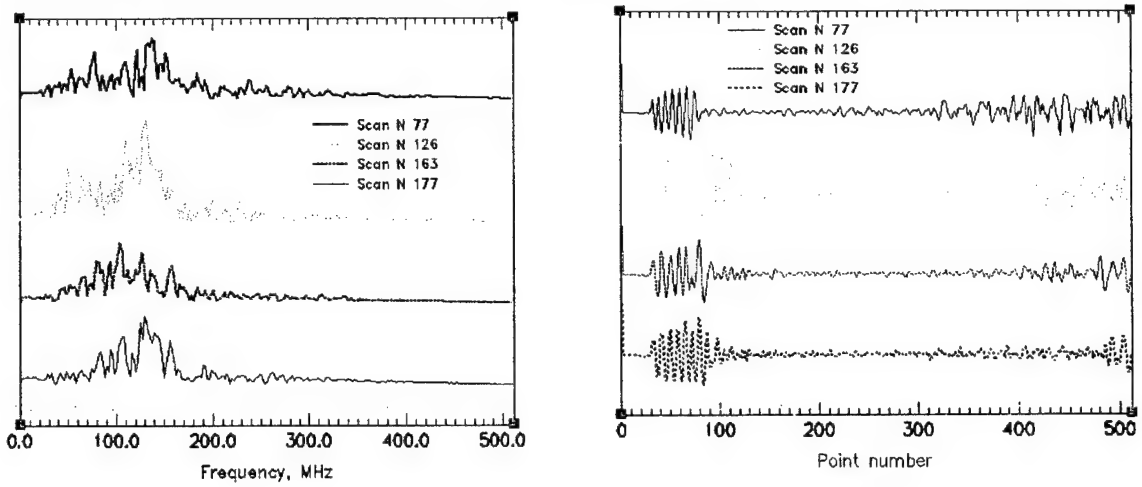


Figure 4-2. Fourier Transform (left) for four A-scans (right).

The most serious argument against the Fourier transform based approach is the similarity of the Fourier spectra for the simulated A-scans with different reflection patterns as seen in Figure 4-3.

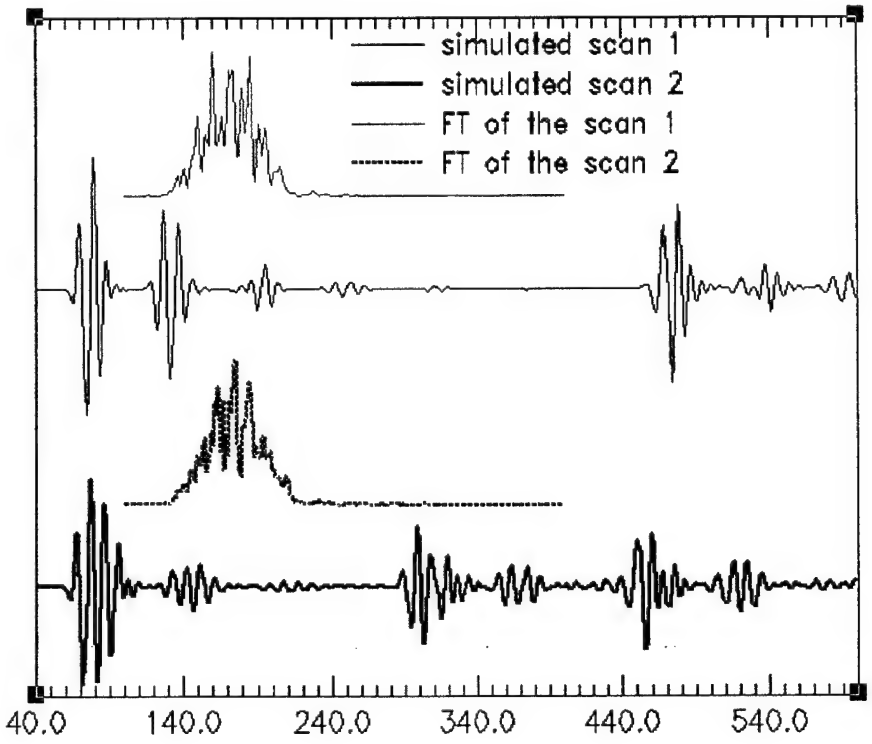


Figure 4-3. Simulated A-scans and their Fourier spectra.

4.4 Wavelet Transform

Wavelet transform (WT) [12] based preprocessing finds extensive application in different pattern recognition tasks [10] and some models of wavelet neural networks were proposed and tested [11]. WT has several advantageous features over the Fourier transform in a sense that WT represents the signal preserving the local neighborhood relations. WT is more flexible in terms of choosing the wavelet basis functions (Figure 4-4) for better data representation.

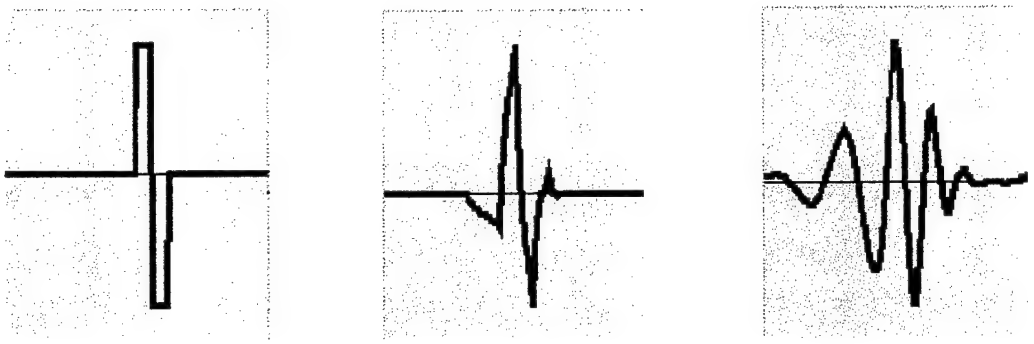


Figure 4-4. Different types of WT basis: Haar, Daubeshies 4, Daubeshies 16.

A conventional Wavelet Transform is a decomposition of a signal into a set of orthogonal basis functions performed in a certain order.

$$A(t) = \sum_{i=1}^{N} w_i \varphi_i^k(t) \tag{4.1}$$

where $A(t)$ is the A-scan in the time domain, w is the weight of the WT decomposition, φ is the corresponding basis function, and k may be considered as a discretization level. The shift and dilation values are usually fixed and on every next step of decomposition are twice as small as on the previous one. Typical A-scan WT decomposition and consequent restoration as a function of the number of WT coefficients is shown in Figure 4-5.

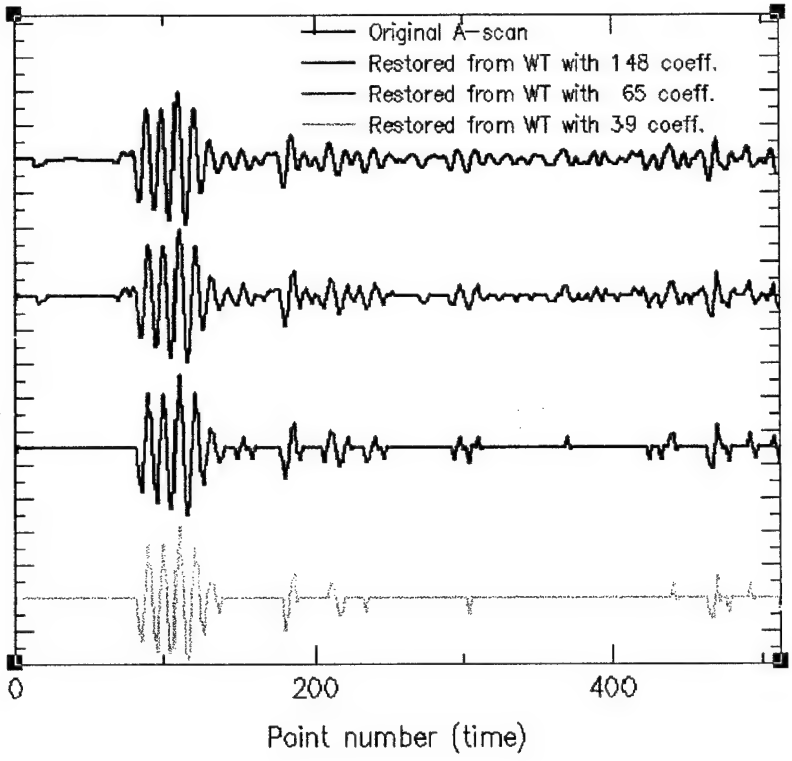


Figure 4-5. Restoration of an a-scan with fewer WT coefficients.

In principle the feature vector for the neural network input may be constructed from the WT components with the largest magnitudes, but the difficulty in this case lies in the inability to predict where (among 512 coefficients) those large magnitudes can appear for an arbitrary scan. That is why the idea of sampling certain groups of coefficients came into play. This technique was implemented and groups of the above mentioned WT coefficients were sampled for different radar lines. Moreover, one could observe the difference between the final distributions of the sampled values of those groups, which were appropriate to be used to distinguish one pattern from another. Unfortunately, after processing a data set of feature vectors constructed as described above no acceptable pattern recognition results were obtained.

The most probable reason was that the approach had a serious disadvantage due to the fixed values of shifts and dilations of this transform - different number and locations of the WT coefficients could occur for the slightly shifted in time refection pattern. Thus, for some situations WT could produce relatively large amount of

redundant data increasing the dimension of the feature vector and ceasing to pick the most discriminatory features of the GPR data.

The next logical step in improving the transform performance was to allow arbitrary shifts and dilations, but this lead to the loss of basis functions orthogonality and prevented the application of the regular WT technique for the signal decomposition procedure. One alternative was the Adaptive Wavelet Transform [10] that gave more freedom to shift and dilation values. But the idea of using as much prior information about the data as possible lead one to reject the wavelet basis and instead choose the best for this specific problem transform basis, which turned out to be the initial pulse used in GPR data collection.

All those issues were major reasons for developing a technique similar to WT for data adaptive signal decomposition called an Adaptive Transform.

Recall that the discussion addresses the default GPR setup: a 600 ns A-scan recorded with a center frequency of 100 MHz, and the term "soil" refers mostly to sands and gravels, unless stated otherwise.

4.5 Adaptive Transform

Adaptive transform (AT) decomposes the A-scan $F(t)$ into a set of initial pulses:

$$G(t) = \sum_{i=1}^{K} \left(w_i^1 I_i(\frac{t-c_i}{s_1}) + w_i^2 I_i(\frac{t-c_i}{s_2}) + \ldots + w_i^l I_i(\frac{t-c_i}{s_l}) \right), \qquad (4.2)$$

where K is the total number of functions in the decomposition, w_i are the weight values, $I_i(t)$ - initial pulse (IP), c_i are the shifts, and s_i are the dilation values. Different values for s are due to the dispersion effect and every next is greater than the previous one (s_1 = 1.0). According to numerical experiments for realistic GPR scan times (500 - 1000 ns) and the actual subsurface media the dispersion is not greater than 1.2. Thus, as few as three or four different dilation values may be used for the decomposition (i.e. s = 1.0, 1.1, 1.2). For the case when we do not account for dispersion (see section 2.3 for the discussion about the applicability of the non-dispersive model) we end up with a

simplified decomposition since $s = 1.0$ and a weight vector with the total number of components equal to the number of sampling points in the signal - K.

$$G(t) = \sum_{i=1}^{K} w_i I_i\left(\frac{t - c_i}{s}\right) \tag{4.3}$$

The residual $|F(t)-G(t)|$ doesn't have to be close to zero (some of the details or noise may be disregarded due to the insignificant influence). There is no need to reconstruct the scan back from the AT. Only the transform coefficients (w, c, s) are used for the data processing. Basis functions of the AT are not orthogonal, that is why decomposition cannot be done in a way similar to Fourier or Wavelet Transform, but it can be performed with the help of the following algorithms.

4.5.1 Correlation-based method (Algorithm 1)

Algorithm 1:

1. find the location t_j of the largest correlation or anticorrelation of the IP and the A-scan:

$$c = \sum_{k}^{POINTS\_in\_IP} I_j(t_k) \cdot F(t_k); \tag{4.4}$$

2. determine the sign and the weight value for a single basis function at this location:

$$w_j = \int F_j(t) I_j(t) dt \Big/ \int I_j^{\,2}(t) dt; \tag{4.5}$$

3. subtract weighted IP at t_j from the original A-scan $F_1(t) = F(t) - w_j I_j(t)$;
4. go to step 1 for $F_1(t)$;

Algorithm 1 in its pure form is ideal for data with relatively low noise level and a low degree of possible overlapping of the reflection pulses. The first problem refers to Equation (4.4) - high noise may result in the largest correlation value to be shifted from the actual position, where it should occur, or in the change of the sign for the correlation value. Due to the single carrier frequency, the high positive correlation in the presence of noise may be misrecognized as a 1/2 period shifted negative correlation (in a similar sense as the $sin(x)$ may be considered to be a shifted $-sin(x)$). The second issue introduces some error into the weight value determined by Equation (4.5) - overlapping

may cause certain changes in the integral value in the numerator: the larger is the degree of overlapping, the less accurate the weight value.

These two problems impose constraints onto the applicability of this algorithm for synthetic as well as real data. Noise issue is absent in the synthetic data, but the Noise Ratio (NR) for real data may vary from a few percent at the beginning of the scan to 100-200% at the last half of it. The most of the reliable GPR information is contained in the first half of the scan (reflections from the top of permafrost and water table), and upon thorough investigation of the underlying physics of the EM wave propagation through the soils some conclusions on the noise nature and its statistical distribution can be made and some elaborate filtering techniques may be applied. Overlapping problems impose limitations onto the minimum spacing between the subsurface layers that the algorithm can resolve with high accuracy. The estimation for this minimum distance for the least possible ε value of about 5 for the permafrost is about 1.3 meters for the IP on the order of 30 ns for a 100 MHz frequency. The constraints and limitations, that were discussed above, do not invalidate the AT concept study and the entire system performance evaluation, since even without the effective filtering technique the neural network processing part has a remarkably high noise and error tolerance (see Chapter 5). Moreover, the alternative method to determine the weight values (though at a higher computational cost), which is discussed in the next section, does not seem to be superior to the correlation-based method at least according to the experiments performed.

4.5.2 Optimization-based method (Algorithm 2)

Algorithm 2 is based on a common optimization problem: we can define an error function for the *F(t)* - the actual scan and *G(t)* - the reconstructed signal from the AT coefficients:

$$E = \sum_{i=1}^{K} \left(G(t_i) - F(t_i) \right)^2 \tag{4.6}$$

over the K sampling points of the signal. Then, using one of the modern optimization techniques for E in Equation (4.6) as a cost function to be minimized over the set of

variable parameters $\{w_1,..,w_K\}$, we can obtain a more accurate decomposition of the scan signal. Logical choice for the optimization technique is a Conjugate Gradient (CG) method [13]. For faster convergence a decomposition generated with *Algorithm 1* may be used as an "initial guess". The major drawback of this algorithm, even with optimal implementation in terms of CPU time, is inability to use this method for possible real-time processing system.

The numerical experiments with elaborate CG software developed by Sergey Perepelitsa [22] showed, that even for a significant degree of overlapping (more than half of the reflection pulse), the decomposition obtained from *Algorithm 1* cannot be improved in terms of adjusting the weight values with the fixed shifts more than by a fraction of one percent. This can be most likely explained by the fact that AT basis function has a length of 30 - 65 sample points (the entire signal is 512 sample points), and changing of one weight coefficient affects the restored scan not only in the region of the CG current improvement, but also in the adjacent region. Thus, very likely, this problem requires some stochastic optimization methods, but those methods are not considered in this work due to extremely high computational complexity, so that real-time implementation becomes infeasible.

4.5.3 Benefits of the AT

This approach has the benefit of very low dimensional signal representation due to the necessity of having only one $I_i(t)$ to represent a single reflection and the property of keeping almost all the possible information about the reflection profile: w is the magnitude and inversion/no inversion information of the reflection coefficient, c is the location in time, s is the relative dispersion. The proposed technique proved to be accurate for the simulated plane-wave formulation data even with thin layers (that produced overlapping reflection pulses within the limits discussed in the Section 4.5.2) and for the simulation data with artificially introduced uniform random noise up to 10% of the average over nonzero components of the initial A-scan (see Figure 4-6). The only

assumption for the AT application is the prior knowledge of the initial pulse propagating through the subsurface medium.

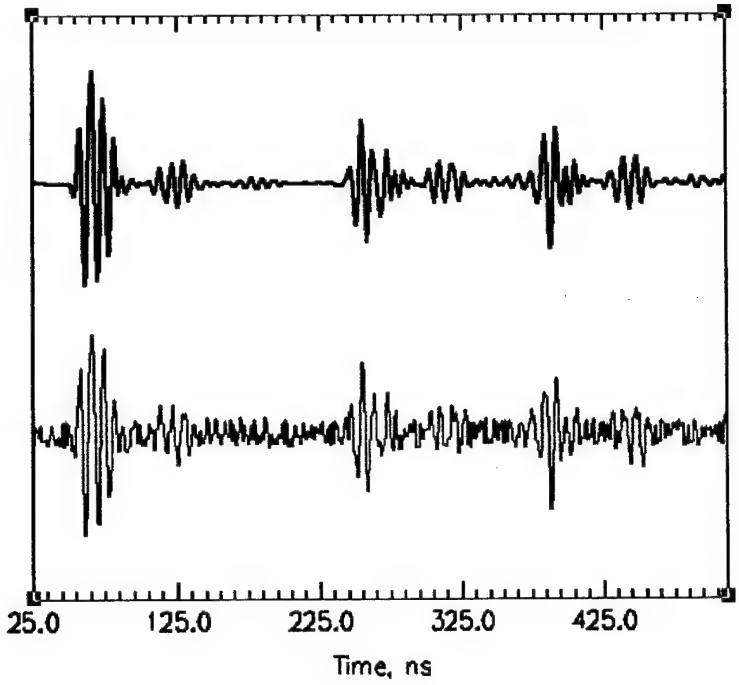


Figure 4-6. Original (above) and noisy (below) synthetic scans.

Important issue is that the initial pulse may be represented not necessarily as an analytical function but may as well be a numerically represented impulse. This makes the AT approach applicable either for synthetic or for real GPR data as far as in the latter case the IP may be determined through identifying especially clear reflection pulse and scaling it to the appropriate amplitude (see Section 4.6.2).

4.6 Adaptive Transform for synthetic and real data

4.6.1 Performance for simulated data

The Adaptive Transform technique was successfully applied to synthetic data: Figure 4-7 shows relative magnitudes w of the weights in the corresponding locations. This demonstrated the capability of the technique to distinguish not only individual reflected pulses but also superimposed ones - w_1 and w_3 are the two reflections from a very close interfaces (1.5 meters thick layer with $\varepsilon = 5.0$). The initial pulse of the form

$$IP(t) = -\cos(\omega t/3)\cos(\omega t) \tag{4.7}$$

was used in the FDTD simulator and in the AT.

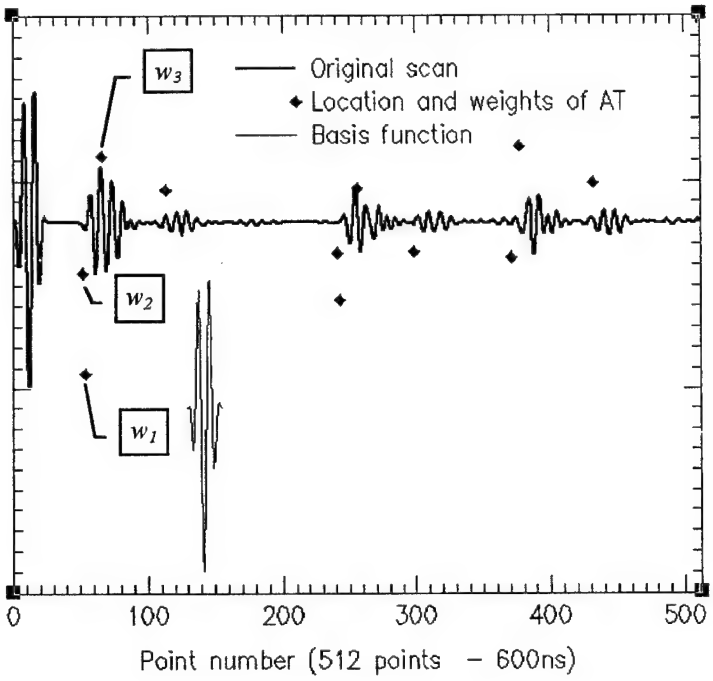


Figure 4-7. Adaptive Transform weights and shifts.

The AT performance results for synthetic scans are neither very interesting nor important because any synthetic scan can be constructed analytically from the reflected initial pulses. With the known geometry and known dielectric constants the scan can be represented as a superposition of reflections from the corresponding layers. Really important results that prove the applicability of the Adaptive Transform are for the decomposition of the real GPR A-scans, which are discussed in the nest section.

4.6.2 Performance for real data

In order to approximate the Initial Pulse a relatively clean reflection from the radar line CR94-61r was extracted. Mathematica 3.0 package was used to approximate the IP by the trigonometric fitting routine - up to 20 terms were used to get an accurate representation. Figure 4-8 shows a plot of this reflection. The scaled version could be considered as a real Initial Pulse used in GPR measurements. Also this representation of the real IP is suitable for embedding into the FDTD simulation model.

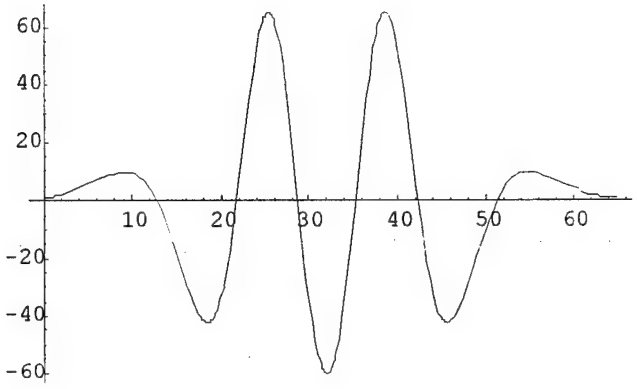


Figure 4-8. Real Initial Pulse.

Application of the Adaptive Transform to real GPR data was verified using the scans from radar lines CR93-11 and CR94-61r. Different initial pulses were extracted from relatively clean reflections due to the fact that the lines have different duration of the scans (600 and 1000 ns) and different number of sampling points (512 and 1024).

Adaptive decomposition was performed with the correlation based algorithm and the scans were restored from the AT shift and weight coefficients to check the accuracy of the technique. For both cases fewer than 10 Adaptive Transform basis functions were used. Figures 4-9 and 4-10 shows the original and the restored scans. A good agreement is observed, major reflections are fairly accurate. Another remarkable property of the Adaptive Transform is demonstrated: noise in the signal is almost completely ignored.

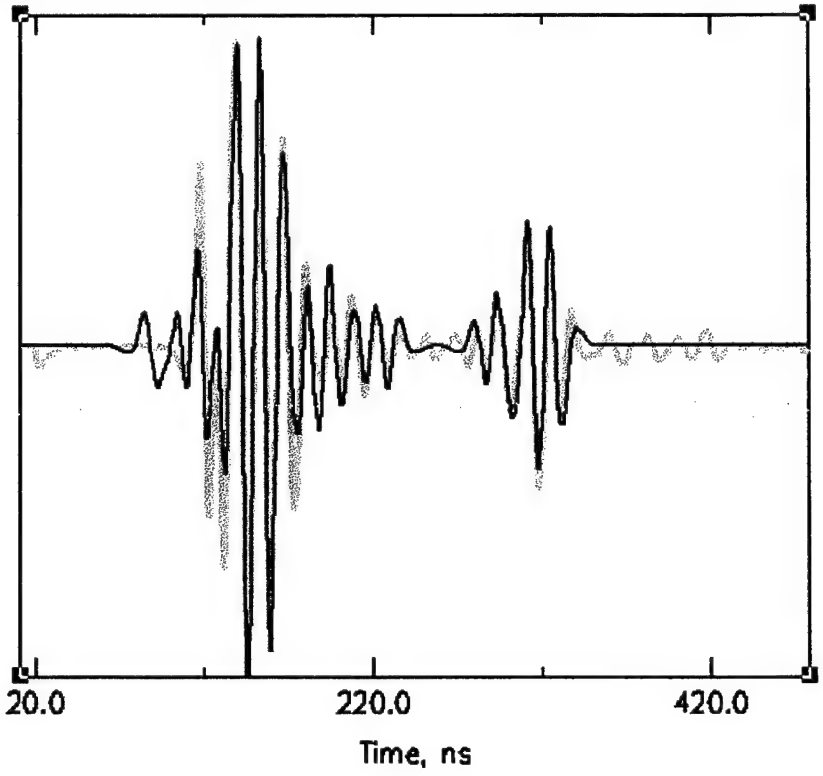


Figure 4-9. Original (gray) and restored from AT (black) scans from CR94-61r.

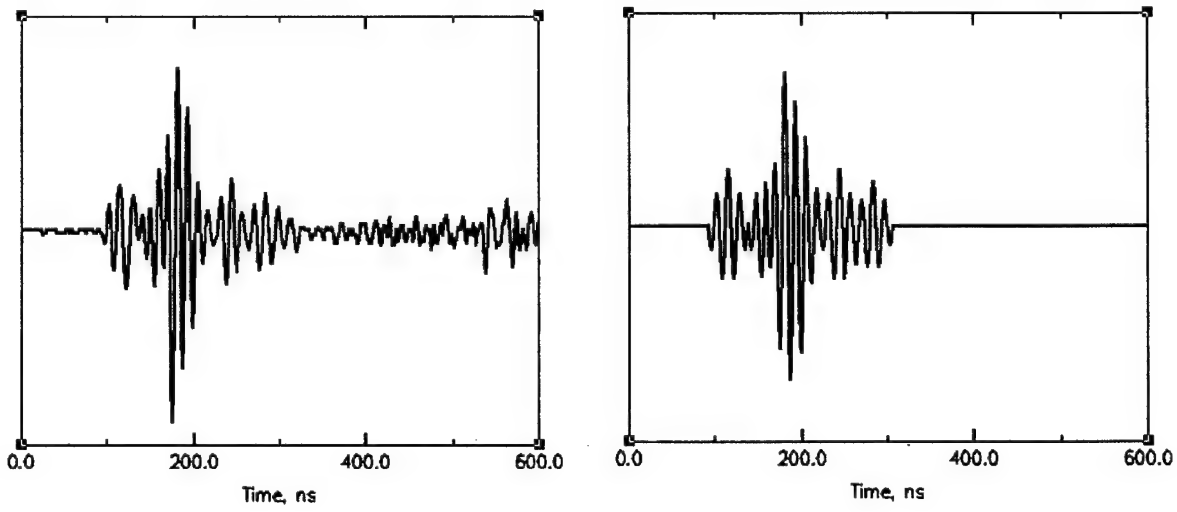


Figure 4-10. Original (left) and restored from AT (right) scan from CR93-11.

4.7 Feature vector from Adaptive Transform

A low dimensional feature vector was constructed from the AT coefficients and used as an input for the neural network processing units. Due to the different formulations (classification and approximation) of NN1 and NN2, different features of the set of AT coefficients were chosen for data representation.

For the classification oriented neural network the first 10 largest in magnitude weights w_i in chronological order were chosen, then the same weights were arranged in descending order of their magnitude and appended to the chronological set. Shift values were not used directly, only the order of the particular weight appearance had importance.

The implications for such a choice of the feature vector are the following: 1) the layers of the single subsurface pattern may be located at different depths, 2) particular time of the reflection is not useful, but the chronological order of the weights and their values that contain the information about the sign of the reflection (whether it is inverted or not) is useful and 3) its strength certainly characterize some subsurface layers configuration. The same weights in the order of magnitude provide some additional information that may become necessary when, for example, a wrong, but not very large coefficient is identified by the AT. In this case the chronological order of weights is disturbed, but the magnitude order remains the same allowing to perform correct pattern recognition.

The approximation oriented NN2 does not require the weight values for the input since the subsurface pattern is already identified and the number and order of layers is known. Now, the exact shift values contain the essential information about the location of the particular reflections and 10 of them that correspond to the first 10 weights exceeding a certain threshold value are used to construct the feature vector.

The construction of the feature vector from Adaptive Transform coefficients is not unique. However, the training sets generated using the above techniques applied to NN1 and NN2, demonstrated very good performance (see Sections 5.3 and 5.4).

4.8 Compensation for variable gain in real system

Another issue, which may be considered as a pre-processing stage is the appropriate normalization of the A-scans. One of the ways to normalize the scan, which is to divide all the values by the maximum amplitude value among all the reflections, was not successful. In this case the strongest reflections for each of the two different A-scans will have the same unit value. Another way, which also does not work well, is the normalization by the average over all points of the scan value - this approach fails due to the different total energy of reflections for different scans.

Our most successful technique to normalize the A-scan by dividing all the components by the maximum amplitude of the initial pulse radiated by the GPR antenna. In this case all reflections will be scaled appropriately - smaller antenna pulses produce smaller reflections, and the relative magnitudes are preserved. Unfortunately, the information about the initial antenna pulse amplitude was not available at the time this work was done. That is why the problem of implementation of the correct normalization technique remains open.

5. HIERARCHICAL GPR DATA PROCESSING SYSTEM

The governing idea of the hierarchical architecture is to split the task of startigraphic layers profiling into consecutive stages, where each subsequent stage has less degree of uncertainty than the previous one. The degree of uncertainty is interpreted in two ways, it becomes larger:

- if the number of the different possible outputs for an arbitrary input;
- if the dimension of the output signal of a current processing stage becomes larger, it makes more difficult to visualize and interpret the results;

The second issue is almost as important as the first one, especially during the development and testing period: the ability of the fast and revealing intermediate result evaluation allows to make the necessary changes in the system in case of bad datum or another system confusion.

The entire GPR signal processing is divided into four pre-processing and two processing stages. Neural network units are chosen to be the main processing blocks due to the following remarkable properties:

- fast operation in the "run" mode after the training procedure has been completed. That permits consideration of real-time implementation;
- high noise/error tolerance [16] which accounts for handling of partially missing or incorrect data.

Data flow and the responsibilities of the particular processing units as well as the operation of the entire system are discussed in this chapter.

5.1 Flowchart

Figure 5-1 shows the detailed data flow diagram and interactions between the separate components. Raw GPR data are used by two pre-processing parts. Pre-processing 0 (see Section 4.1) checks the data consistency and excludes inadequate scans from the regular processing. Pre-processing 1 incorporates some information about the previously processed scans and is described in Section 5.2.

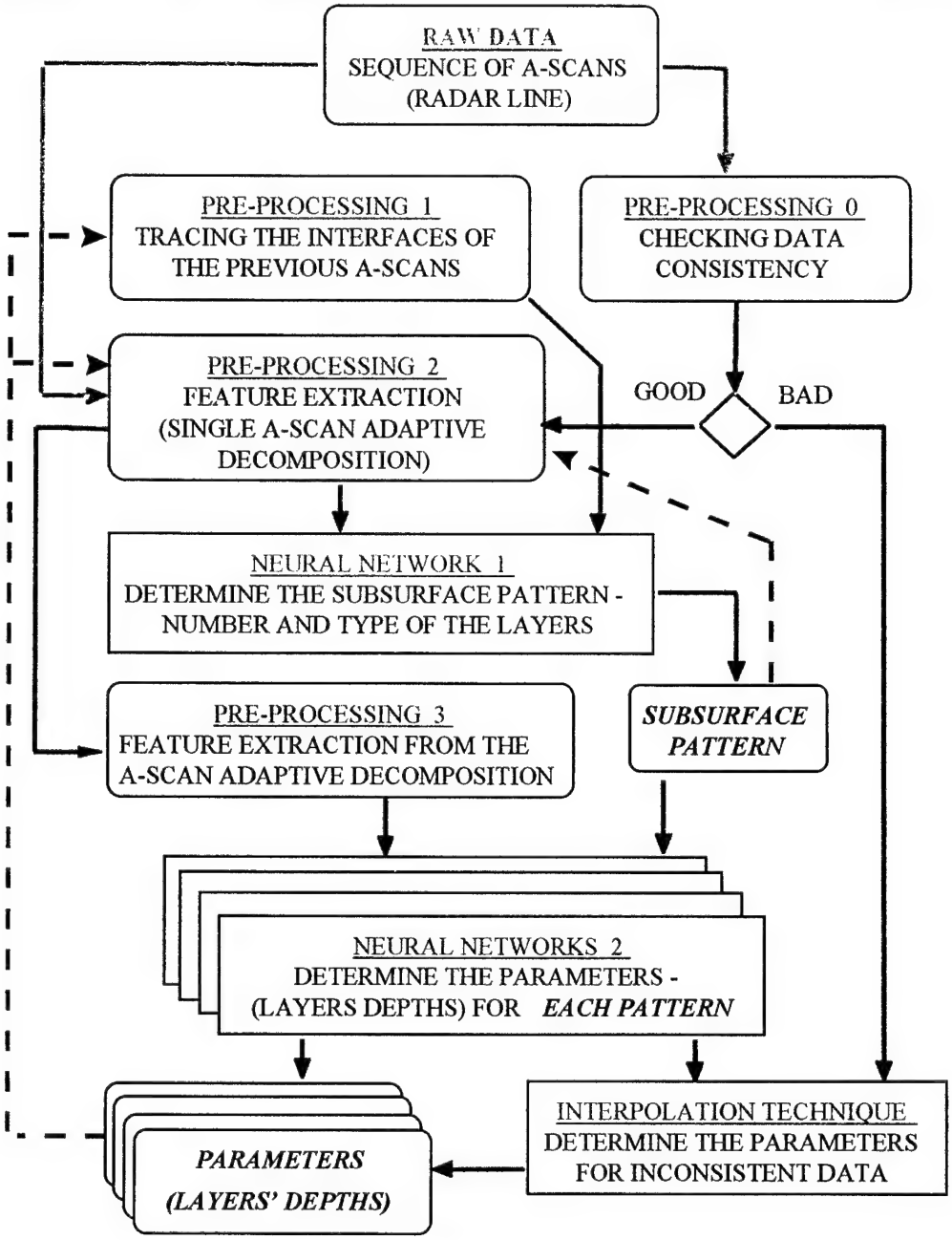


Figure 5-1. Complete algorithm flowchart.

Pre-processing 2, a critical component, decomposes the A-scans into the small feature vector which containe essential and highly discriminative information about the data (Section 4.5). The first main processing unit Neural Network 1 (NN1) is used to recognize the subsurface pattern of the current scan (Section 5.3). Pre-processing 3

(section 5.4.1) extracts some other useful feature from the adaptively decomposed GPR scan and transforms the result to fit the Neural Network 2 (NN2), which by combining this information and the knowledge of the current subsurface pattern determines the depths of the stratigraphic layers (Section 5.4). A separate NN2 module is required for each subsurface pattern.

Several interpolation techniques can be used to determine the layer depths for the inconsistent data. If the number of "bad" scans is significant – an investigation of the soil, hardware or control software adjustment is warranted.

Feedback possibilities in this processing algorithm are shown with dashed lines. The one which goes from the Layer Depths to the Pre-processing 1 provides the history of the depths values to be able to predict ones for the current datum. The other two from Subsurface Pattern and the Layer Depths to the Pre-processing 1 allow to rerun the entire algorithm for the particular scan if it appears to be recognized incorrectly or its parameters determined inaccurately. Then the necessary modifications are made in the process of adaptive decomposition and the new feature vector is generated. Possible implementation of those techniques are discussed in the Sections 5.3.7 and 5.4.5 of this Chapter.

5.2 *Tracing interfaces depths technique*

This technique (pre-processing 1 on the block diagram) is based on the physical properties of the subsurface media in the region of interest - the actual geographic area. Those properties imply relatively flat water table boundary (that is quite natural for any water surface) and relatively flat top of the permafrost - based on the actual observations. On the one hand, the approach provides a statistical curve fitting technique to predict the location of the particular interface for the current A-scan, based on the information about several previous scans. On the other hand, the tracing procedure is used for creating of the top-down "depth-based" feedback (Section 5.4.5). If the particular interface depth, identified with NN2 unit, for the current scan differs significantly from the corresponding depth for the same interface determined by the statistical technique - an alarm signal is sent to the Adaptive Transform stage to pay more attention to the particular time interval

that has caused the mismatch and the necessary adjustments for some of the feature vector components are to be made. The implications of this approach are similar to the topology based or context based techniques in Optical Character Recognition (OCR) applied to handwriting or poorly typed printed texts: severely distorted single character can hardly be recognized by the OCR system, but some of its resolved features coupled with the information about neighboring characters (like continuos topology for handwriting or reasonable context that produces a word with a common meaning) may end up with correct final recognition.

In this system the technique of tracing the interfaces depths was implemented by means of linear regression models [15]. Relatively flat interfaces (like water table) require only linear term (model 1 - equations 5.1 - 5.3), and the interfaces that exhibit actually large variations (bottom permafrost boundary) may be predicted with the model with the x^2 term present (model 2 - equations 5.4 - 5.7).

$$d(x) = y = a + b \cdot x \tag{5.1}$$

$$b = \frac{\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i \right) / n}{\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i \right)^2 / n} \tag{5.2}$$

$$a = \overline{y} - b \cdot \overline{x} \tag{5.3}$$

where the bar over the terms means simple averaging, and n is the number of points used for fitting the interface depth $d(x)$, and x is the normalized distance along the radar line.

$$d(x) = y = a + b_1 \cdot x + b_2 \cdot x^2 \tag{5.4}$$

$$b_1 = \frac{\sum_{i=1}^{n} x_i y_i \sum_{i=1}^{n} x_i^4 - \sum_{i=1}^{n} x_i^2 y_i \sum_{i=1}^{n} x_i^3}{\sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} x_i^4 - \left(\sum_{i=1}^{n} x_i^3 \right)^2} \tag{5.5}$$

$$b_2 = \frac{\sum_{i=1}^{n} x_i^2 y_i \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} x_i y_i \sum_{i=1}^{n} x_i^3}{\sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} x_i^4 - \left(\sum_{i=1}^{n} x_i^3\right)^2} \tag{5.6}$$

$$a = \bar{y} - b_1 \cdot \bar{x} - b_2 \cdot \overline{x^2} \tag{5.7}$$

The number of points (nodes) to be used for fitting and the spacing between them is user defined. Four nodes are generally used and the total distance between them does not exceed 4 meters. Figure 5-2 illustrates the prediction of the interface depths with model 1 and model 2, respectively.

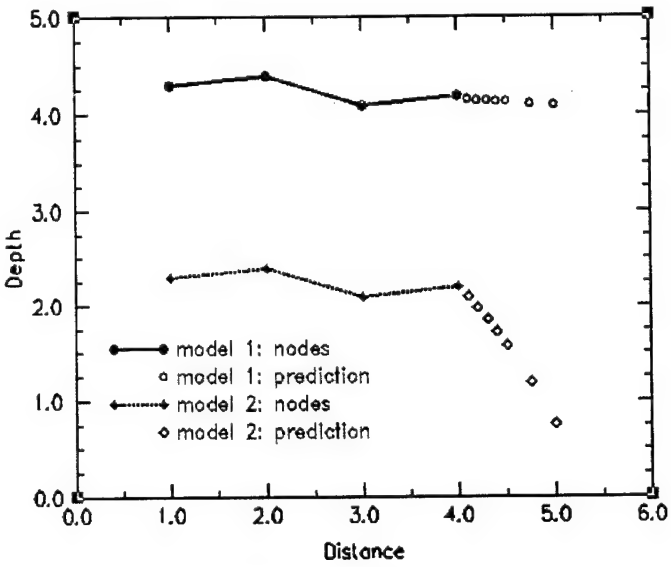


Figure 5-2. Regression models performance.

5.3 *Subsurface pattern identification*

This section covers the process of subsurface pattern identification starting with the rules for classification of the GPR data into categories, encoding, followed by the description of the Neural Network 1 architecture, parameter choice, process of generating a training set from the Adaptive Transform data, and training itself.

5.3.1 Classification and encoding of the subsurface patterns

For the experiments with synthetic data seven subsurface patterns were defined and subsequently used for the training sets: the example of the pattern is presented in Table 5.1, a complete set of subsurface patterns is provided in Appendix B. Information about the season is highly important when assigning the appropriate ε values to the different layers. Layers, that may be frozen in spring, may be thawed in autumn,

changing significantly the reflection pattern at the beginning of the scan as well as the reflection times for the deeper layers.

Table 5.1. Example of the subsurface pattern.

Pattern Number	Season	Layers	Dielectric constant
1	Autumn	Active layer (organics)	14 - 25
		Silt or sand (dry)	< 20
		Sand or gravels (saturated)	20 - 45
		Permafrost	4.4 - 5.6

There is only one rule for classification of the possible combinations of the stratigraphic layers into subsurface patterns - each new combination of the existing layers or appearance of any additional layer generally leads to a new pattern.

Patterns are encoded using a 1-OF-C (binary encoding for each category) [20] encoding method. Number of outputs of the Neural Network 1 is equal to the number of patterns used for this particular simulation:

Table 5.2. 1-OF-C patterns encoding.

	Output 1	Output 2	Output 3
Pattern 1	1.0	0.0	0.0
Pattern 2	0.0	1.0	0.0
Pattern 3	0.0	0.0	1.0

5.3.2 Generation of the training set

The training set was automatically generated from the Adaptive Transform coefficients: shift and weight values. Information about the patterns structure (number of patterns, number of inputs and outputs) was stored in the training set as well as the patterns itself (please note, that here the term "pattern" does not mean the "subsurface pattern", it only means a single training example - one feature vector, that corresponds to a single GPR A-scan). A simple technique was used to construct the input (and nevertheless proved to be accurate enough for correct processing): 10 AT weight values

in chronological order followed by the same 10 weights in the descending order with respect to their magnitudes. Implications for this particular choice for the feature vector construction were discussed in Section 4.6. This feature vector choice might not be the optimal one, but it was the best at the time this work was done.

5.3.3 Neural Network 1 architecture and parameters

The block diagram of the Neural Network 1 operation is shown in the Figure 5-3.

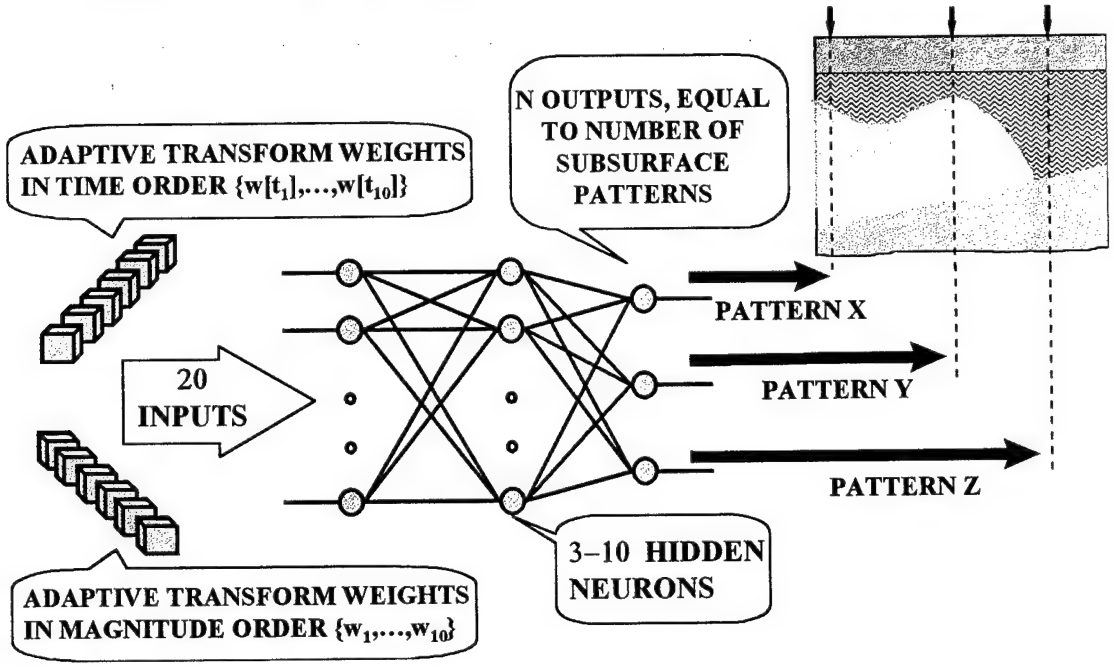


Figure 5-3. Neural Network 1 architecture.

The architecture used was an ordinary feedforward layered neural network - a Multilayer Perceptron with a single hidden layer. Number of neurons in the input and the output layers correspond to the number of inputs and outputs for a single pattern in the training set. One part, which is a user responsibility, is to set the appropriate number of neurons in the hidden layer - there are no absolute rules [20], but certain guidelines address the problem of overfitting [20], [21]. Overfitting occurs when the network is "overtrained" - it can not exhibit a good generalization property due to the fact that it has learned all the training patterns in great details and is not able to perform well on the data it has not seen before (not from the training set). One of the methods that helps avoid

overfitting is to choose the total number of weights (by setting the appropriate number of hidden neurons) in the network to be several times less than the total number of training patterns.

Two different algorithms were used for training of the network: on-line Backpropagation and Scaled Conjugate Gradient (SCG) methods covered in the Chapter 3. Their performance in terms of computational time and resources was not significantly different due to the small size of the problem. Since SCG is more complicated and at present implementation is less suitable for results evaluation and visualization and mostly intended to solve large scale problems, the discussion on performance, noise and error issues is limited to the more conventional Backpropagation method.

Another essential feature of the neural network implementation is data normalization [5]. Several normalization methods are applicable to the MLP networks: column normalization (each input variable normalized separately), full normalization (all variables normalized by the same factor, i.e. the largest input value among all the variables). In this implementation full normalization was employed due to the fact that absolute values of the inputs (AT weights) were of the same order. Column normalization may be helpful for data with a significant variables difference in magnitude.

5.3.4 Neural Network 1 training and performance

The training was performed on a training set generated from the synthetic GPR line based on the first (west) 600 meters of the real radar line CR93-11. Four different subsurface patterns were defined (patterns 1, 2, 3, and 4 in Appendix A). Training set with 450 patterns corresponding to the same number of the artificial A-scans was used.

The network total operation time (including training) for a DEC ALPHA (175 MHz) machine was 10 seconds for the network architecture with 10 hidden neurons, and 7 seconds for 3 hidden neurons. The convergence plot is presented in Figure 5-4. Both networks each the desired error level, but the one with 10 hidden neurons did so in fewer number of iterations.

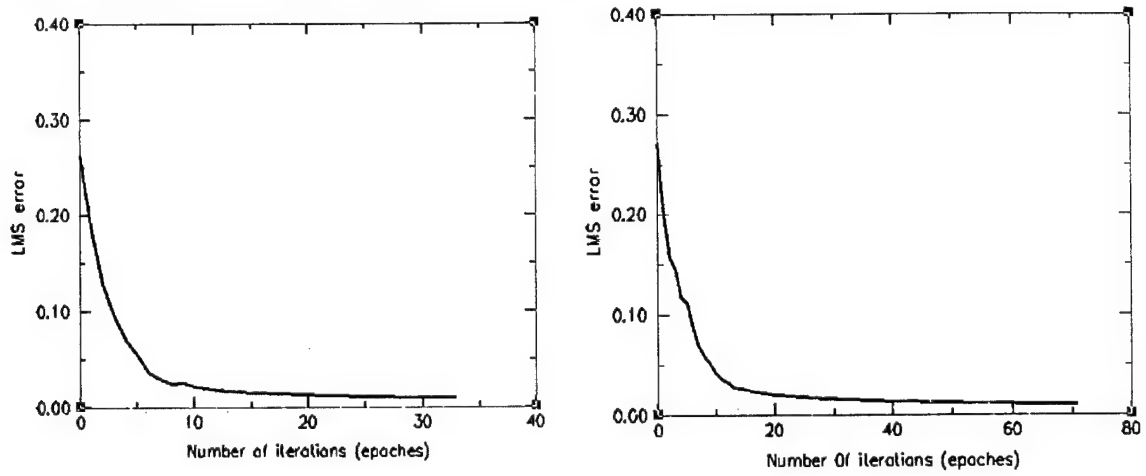


Figure 5-4. Convergence of NN1, 10 hidden neurons (left), 3 hidden neurons (right).

5.3.5 Pattern identification results

A synthetic test set was used to validate the network accuracy and generalization. Thirty artificial scan examples representing all four subsurface configurations, that were not presented to the network as part of the training set were tested - every single input vector was recognized correctly. Network activation values (black) vs. actual desired outputs (gray) are shown in Figure 5-5. Network output was slightly smaller than the actual value it should be, only because of the fact that normalized output values (0.5) and (-0.5) were the maximum and minimum values the network had learned and, due to its properties, it couldn't extrapolate beyond the range of the values, that were presented during the training.

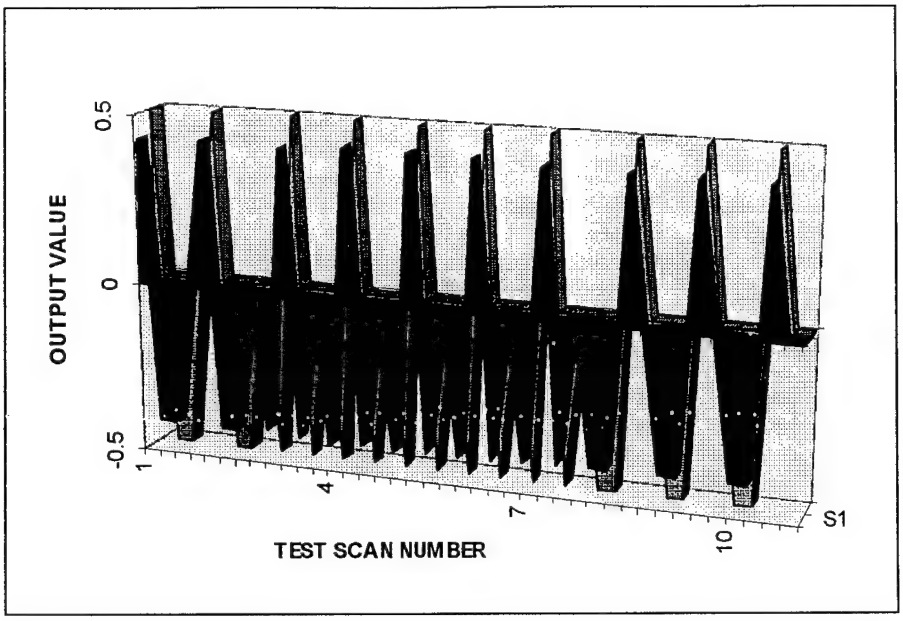


Figure 5-5. Neural Network 1 pattern recognition accuracy.

5.3.6 Testing with real data

Three different A-scans from the GPR line CR93-11 were used to verify the entire processing algorithm performance for the ultimate problem, it was designed for - recognition of real data by the neural network trained with synthetic data. Unfortunately, even with the borehole information available it was not always possible to determine the full information about all the details of the subsurface media profile corresponding to the particular A-scan. That is why the evaluation of the NN1 accuracy for the real data recognition was made based not on a 100% correct information, thus, it has a slightly fuzzy meaning. Table 5.3 gives the example of the tests result for 3 A-scans: scans 1614 and 2700 very likely represent the subsurface configurations defined as Pattern 1 and Pattern 2 respectively, scan 8810 is extracted from the East part of the CR93-11 and, probably, does not correspond to any of subsurface patterns used for this simulations.

The left part of the Table 5.3 shows the activation values of the output neurons as a respond to the feature vector generated with the Adaptive Transform from the actual GPR A-scan. The right part contains the ideal (normalized) responses for the four subsurface patterns used for NN1 training. The comparison results in correct recognition

for the scans 1614 and 2700, which very likely correspond to the actual configuration of the stratigraphic layers similar to the patterns 1 and 2, respectively. Scan 8810 was not classified into any category, and this does not contradict with the assumptions made.

Table 5.3. NN1 recognition results for real data.

Scan number	1614	2700	8810		Pattern 1	Pattern 2	Pattern 3	Pattern 4
Activation of neuron 1	0.922	-0.734	-0.938		1.0	-1.0	-1.0	-1.0
Activation of neuron 2	-0.880	0.875	-0.726		-1.0	1.0	-1.0	-1.0
Activation of neuron 3	-0.917	-0.801	0.628		-1.0	-1.0	1.0	-1.0
Activation of neuron 4	-0.911	-0.932	0.828		-1.0	-1.0	-1.0	1.0

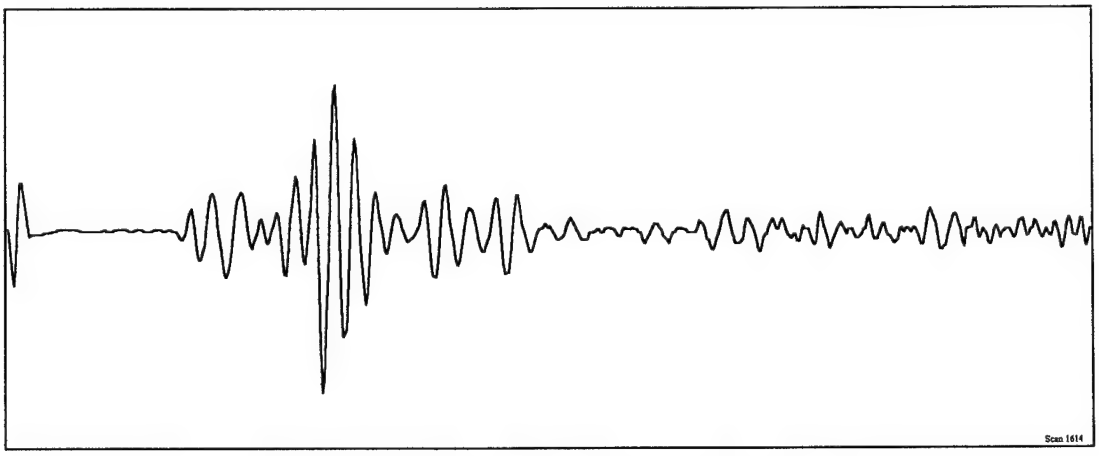


Figure 5-6. A-scan 1614 from GPR line CR93-11.

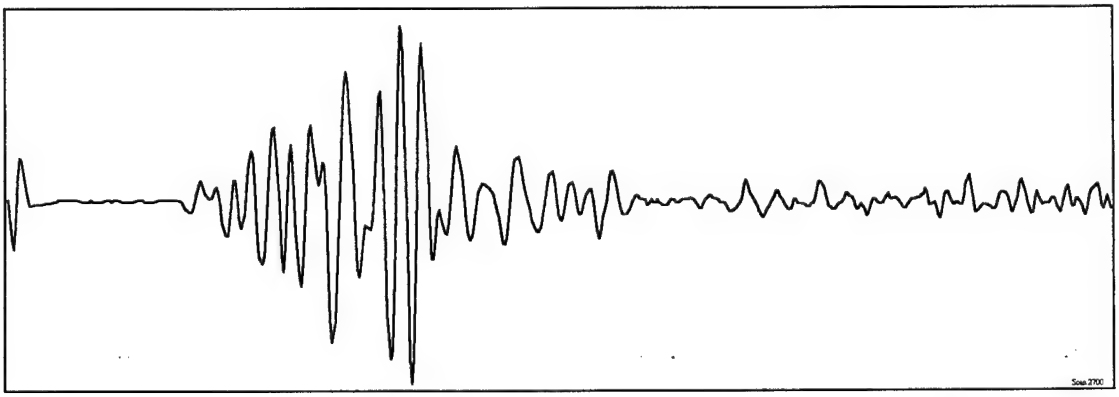


Figure 5-7. A-scan 2700 from GPR line CR93-11.

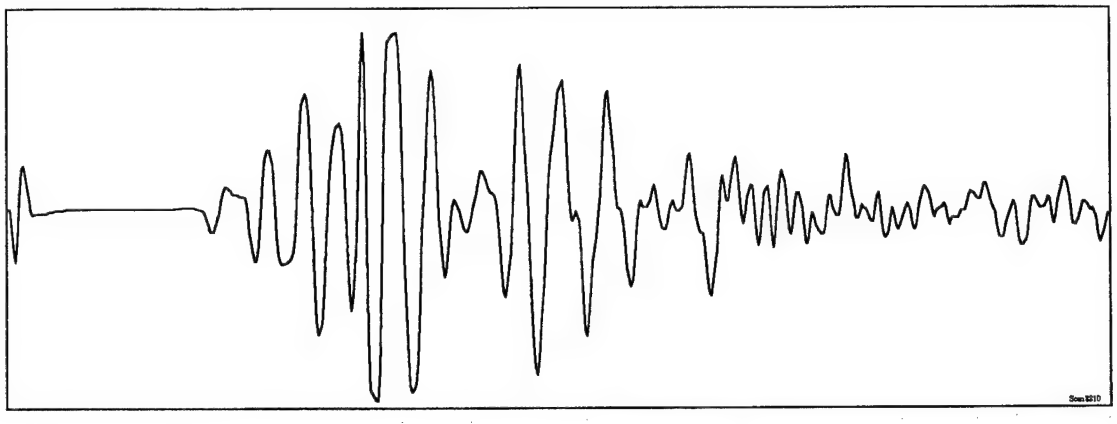


Figure 5-8. A-scan 8810 from GPR line CR93-11.

5.3.7 Pattern-based feedback

Actual GPR line may be composed of the scans that belong to different subsurface patterns, but, in general, the properties of the vertical soil and soil structure (like permafrost) distribution do not undergo significant changes over the distance of meters or tens of meters. Thus, with the collection rate of 5-10 scans per meter, the adjacent scans very likely represent the same subsurface pattern, and occurrence of the other pattern is most probably due to the incorrect pattern recognition.

When the different subsurface pattern occurs the proposed feedback generates an alarm signal and sends it to the Adaptive Transform pre-processing stage. AT addresses the correlation part (Section 4.5.1) and checks for the high correlation values, that did not take part in the scan decomposition because they were suppressed by the other slightly higher correlations nearby. This could be caused by noise or other factors. On the next step a different correlation is chosen suppressing the formerly dominating one and the decomposition is performed again, followed by the new attempt of subsurface pattern recognition by NN1.

This method has not been implemented in software partly because of the lack of availability of some essential GPR hardware related information preventing the development and implementation of the fully automatic data processing systems. Please refer to Chapter 6 for further work discussion.

5.4 Determining depths of the subsurface layers

This section discusses the second part of the main processing architecture, developed for the identification of the depths of the particular layers for the known subsurface pattern. The general idea of this approach is expressed in the Figure 5-9.

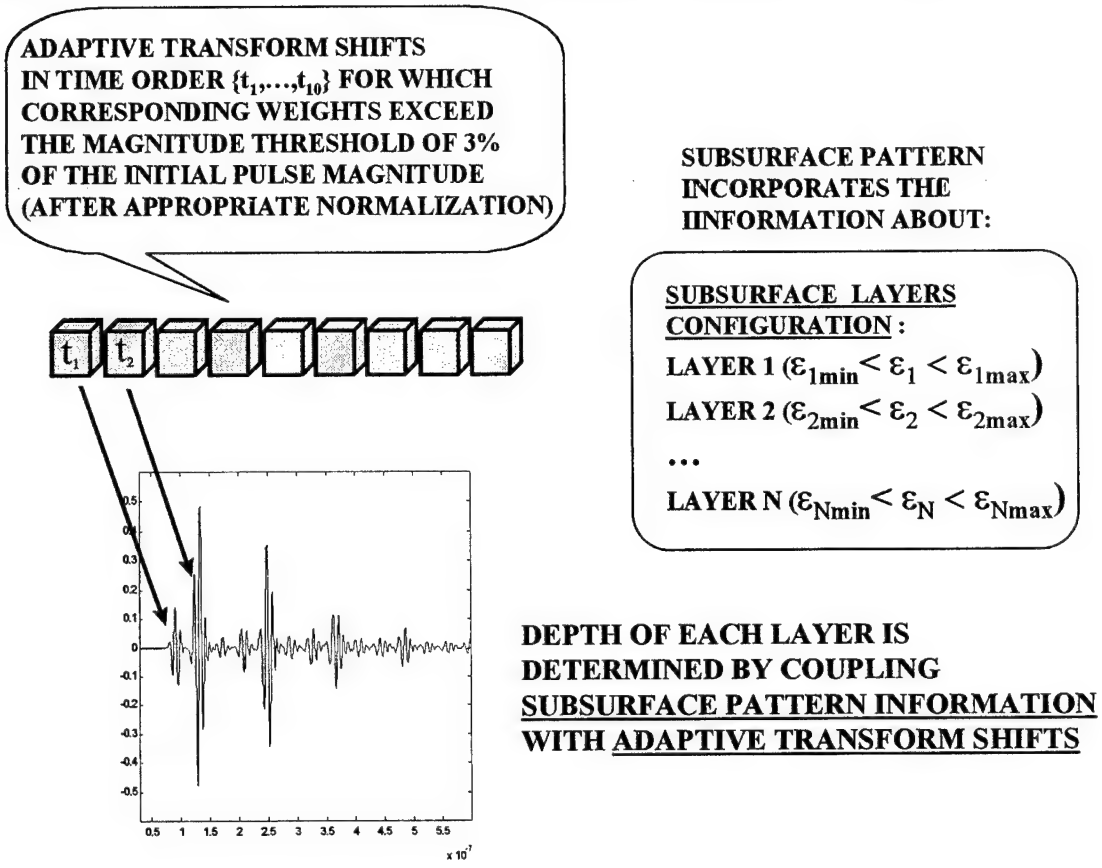


Figure 5-9. Layers depths identification setup.

5.4.1 Generation of the training sets

The training set for Neural Network 2 was automatically generated from the Adaptive Transform coefficients. Another technique was used for the assembling of a single training pattern due to the different from NN1 purpose of this processing stage - the approximation of certain continuous parameters.

First, ten AT shift values which corresponded to the weights with the magnitude exceeding some threshold were chosen and arranged in chronological order. Then, for

each of the 10 variables the average value was calculated. At the next step each variable was replaced by the difference of its original value and the corresponding average. The same operation was performed for the depth values used as an output for the Neural Network 2.

This procedure allowed the incorporation of the essential information about data peculiarities into the input/output variables, as a consequence of the increased range of change in values: substitution of deviations for the actual values. Finally, this technique for generating the feature vectors for the training set improved significantly the accuracy of the depth identification.

5.4.2 Neural Network 2 architecture and parameters

Architecture of the Neural Network 2 is basically the same as of the Neural Network 1 - a Multilayer Perceptron with one hidden layer. The difference is in the number of neurons in the layers - as far as the information about the subsurface pattern is already known, the number of possible parameter distributions (depth values) is smaller, than for the data set containing different subsurface patterns, that is why fewer inputs may be used for correct network operation.

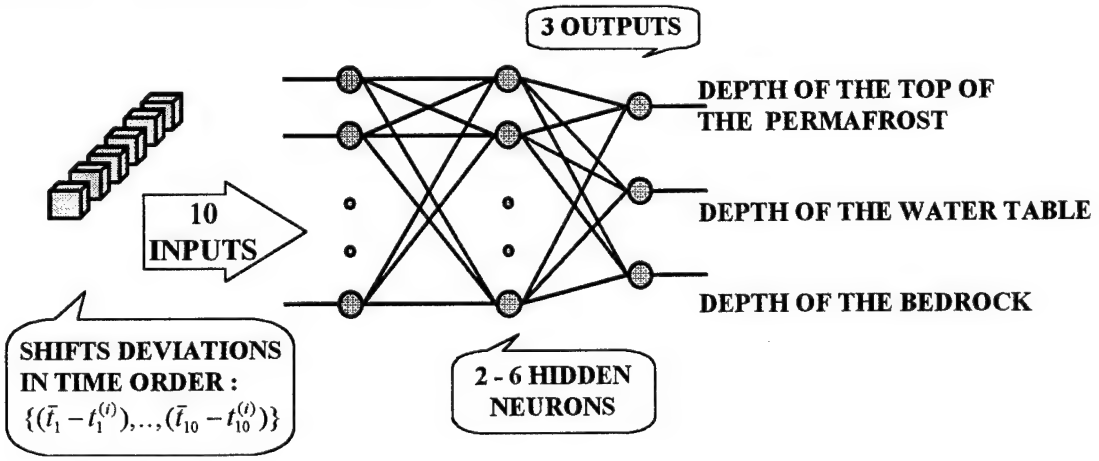


Figure 5-10. Neural Network 2 architecture.

The problem formulation for this processing part is "approximation". The continuous parameters are used as the target outputs for the error correction during the network training instead of binary categories for NN1. But this difference does not affect

the training procedure with the exception of the stopping criterion. The possible overfitting was more critical for NN2 than for NN1 and, besides the appropriate architecture choice, the "early stopping" method [20] is employed. This was accomplished simply by making the stopping training tolerance higher than, for example, the set value for NN1.

5.4.3 Neural Network 2 performance

The network was trained with the data set generated from the same CR93-11 synthetic GPR line, but only the patterns that belong to the subsurface pattern 2 were chosen to be included into the training set. A total of 100 patterns were used, 92 of them for training purposes, 8 for verification of the network performance.

Networks with different number of hidden neurons were tested and the performance in terms of final accuracy was measured (Figure 5-11). The results verify the statements about the appropriate choice of the number of hidden neurons - too many neurons in the hidden layer lead to overfitting and worsen the performance on the testing set, the network generalization and consequently interpolation property becomes poor - the results are less accurate. CPU time (DEC ALPHA 175 MHz, seconds) does not reveal a lot of information because larger networks have more weights participating in computation but it may converge faster due to the ability to store more information.

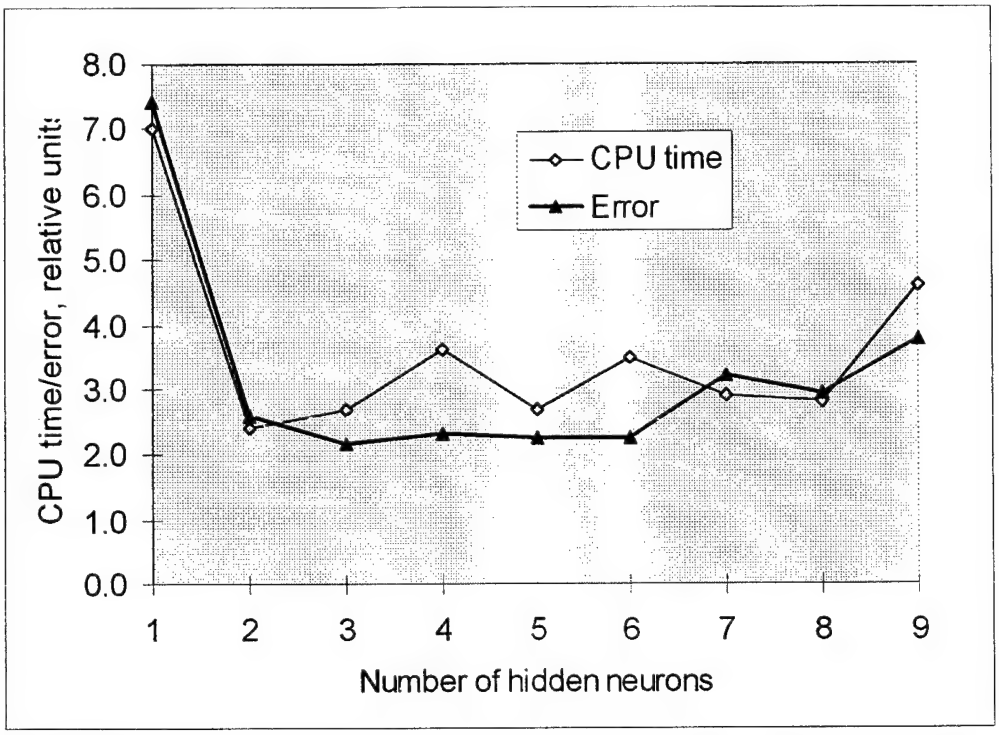


Figure 5-11. NN2 performance for different number of hidden neurons.

The network with 3 hidden neurons was chosen as a final one. The plot of convergence is shown in the Figure 5-12.

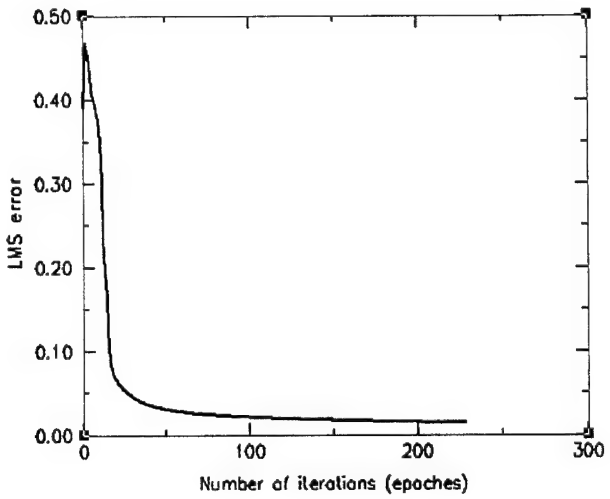


Figure 5-12. NN2 convergence.

5.4.4 Depths determining results

Accurate results for the depths of 3 layers for synthetic GPR line CR93-11 subsurface pattern 2 were generated as an output of NN2. The average error was about 4% for the transformed and normalized data. Due to the discussed above technique for the training set generation and the built-in network data normalization, the error was the same regardless of the absolute depth of the layer. The maximum absolute error for this simulation of about 0.5 meters was produced as a result to approach the boundary of the range of the possible network output values.

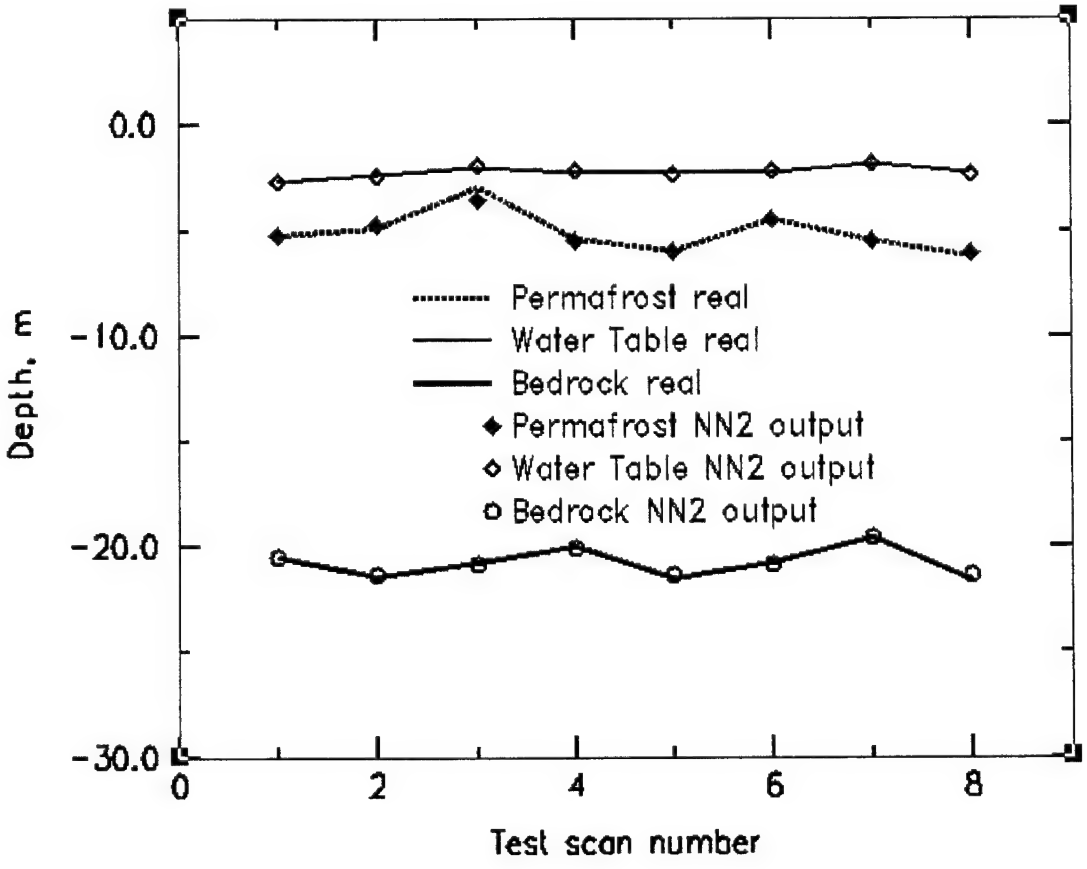


Figure 5-13. Layer depths identification results with NN2.

The results are presented in Figure 5-13, plotted values do not reflect the exact distribution of the layers because in real life depths are measured from the ground surface - in this simulation model they were measured from the imaginary zero level. But the

resulting values may be transformed into stratigraphic information incorporating the data about the surface elevation.

5.4.5 Testing with real data

Data from the same GPR line (CR93-11), as in the Section 5.3.6 were used to verify the applicability of NN2 module trained with synthetic data to real test cases. NN2 was trained to determine the depths of the subsurface layers for the scans that belong to subsurface pattern 2, that is why only the scan 2700 was used. Results are summarized in Table 5.4 below.

Table 5.4. Testing results for NN2 for real data.

Scan number	2700	Depth, predicted by NN2, m	Depth from Borehole data, m	Layer type
Activation of neuron 1	0.231	2.25	2.40	Water table
Activation of neuron 2	0.468	4.08	4.50	Permafrost
Activation of neuron 3	0.699	19.88	25.50	Bedrock

The accuracy may be considered good taking into account the fact, that borehole data are not 100% reliable, and the scan location on the GPR line might not be very close to the hole drilled. Nevertheless, assuming the correctness of the experimental borehole information for this case, the following conclusions can be made. Unlike the situation when the synthetic data used as the test examples, for the real test examples the error increases with depth. There are at least two reasons: the first one accounts for the longer traveling distance of the electromagnetic wave reflected from the distant interface resulting into more scatterers on its way, hence more noise is accumulated in the second half of the A-scan. The second reason is concerned with the possible change of the ε value - the longer is the distance, the more effect it may produce on the reflection time versus layer depth relation.

No extensive testing was performed mostly due to the lack of the reliable and accurate in all three dimensions experimental geophysical information, which may be

matched with the collected real GPR data resulting in reasonable subsurface media configurations .

5.4.6 Depth-based feedback

This feature of the entire processing architecture can be used for two different purposes. One is just to provide the Pre-processing 1 with the current depths values in order to trace the interface depths for the next scan processing. Another goal is identical to the pattern-based feedback discussed in Section 5.3.7: if the depth of a particular interface is determined and is out of some reasonable range, an alarm signal is sent to the AT and the same procedure for picking the different correlation point is performed. The only difference is that the range of search for the different high correlation value is narrowed down to the neighborhood of the reflection corresponding to the interface of the current interest.

5.5 *Accuracy, noise tolerance*

For the hierarchical architecture it is not always possible to evaluate the algorithm accuracy as a single system. Some information passed from one stage to another is binary and the accuracy issue narrows to a decision "correct" or "incorrect". Accuracy of Adaptive Transform, Neural Networks 1 and 2 as the separate processing units is discussed in Sections 4.5, 5.3.6, and 5.4.4, respectively. This section provides the estimation of the system (Adaptive Transform + NN1) accuracy in terms of subsurface pattern recognition sensitivity to the change of the ε value for the synthetic test examples.

Dielectric permittivity value for one or two media were varied and the table below summarizes the results. These results were generated by averaging over 40 test examples.

Table 5.5. Subsurface pattern recognition sensitivity to the change of the dielectric permittivity value.

Number of mediums varied	Difference from the original ε value when recognition becomes incorrect, %
1	30 - 40
2	20 - 25

6. CONCLUSIONS

6.1 Summary

As a result of this research the following separate Ground-Penetrating Radar (GPR) data processing system components have been developed and implemented in software.

- Adaptive Transform (AT) signal decomposition technique, which allows the handling of uncertain and noisy raw GPR data by extracting representative and discriminative features;
- Three additional pre-processing methods, which provide the possibility of eliminating inconsistent data, utilizing the information about previously processed data, and filtering the AT feature vector in an effective way;
- Two neural networks modules that perform major processing functions based on either Backpropagation or Scaled Conjugate Gradient training algorithm.

Performance of the each component is tested with synthetic as well as real GPR data and has demonstrated acceptable operation with respect to accuracy and error/noise tolerance.

The hierarchical GPR data processing system is constructed from the above mentioned components. Consecutive operation of the developed algorithm stages allows to decrease the degree of data uncertainty for each of the following processing step.

The entire system, constructed and trained with synthetic data, is tested with real GPR information. Successful results and robustness of the algorithm have established the feasibility of applying the approach to actual GPR related scientific and technological problems. This research can now become the basis for further development and improvement.

6.2 Further work

Although the feasibility of applying the proposed system is established, certain improvements and additions should be made. The first group of improvements relates to both synthetic and real GPR data. The current model for synthetic data generation may be extended, by including more physical factors when simulating soil and soil structure electrical properties. Factors such as wave attenuation, dispersion, and scattering are the primary goals in this study. Taking into account these aspects, more understanding of the underlying physical processes could be achieved. This may yield a better understanding of the behavior of electromagnetic waves in subsurface media revealing the cause of uncertainties in GPR data related phenomena. Certain improvements in the GPR data collection technology in terms of antenna gain control and availability of additional useful information will lead to overcome such difficulties as inability to correctly normalize the data and define the proper basis functions for the AT.

The second group of improvements is algorithm related. The following ideas were considered, but were not implemented due to the lack of time.

- extension of the AT into a multiresolution algorithm permitting the handling of wave dispersion;
- development of a more efficient way for constructing the feature vector from AT coefficients;
- testing Radial Basis Functions neural network paradigm as an approximation processing tool;
- implementation of a pattern-based and depth-based feedback along with the optimization technique to compensate for possible errors on any processing stage.

The concept of a fully automatic GPR data processing system is a promising engineering area considering the growing demand in environmental, industrial and other applications. Development of a working system, which can operate in conjunction with existing GPR equipment, will certainly be of benefit to our industrial society.

7. REFERENCES

1. Steven A. Arcone, Daniel E. Lawson, Allan J. Delaney, Jeffrey C. Strasser, Jodi D. Strasser, "Multi-bandwidth Reflection Profiling of Discontinuous Permafrost with Ground-Penetrating Radar", CRREL Report , U.S. Army Cold Regions Research and Engineering Laboratory, Hanover, NH, 1997.

2. Smithsonian Physical Tables, 9th edition, Smithsonian Institution, Washington, DC, 1954.

3. E. I. Parkhomenko, "Electrical Properties of Rocks", Plenum Press, NY, 1967.

4. Christopher M. Bishop, "Neural Networks for Pattern Recognition", Clarendon Press, Oxford, 1995.

5. Simon Haykin, "Neural Networks: A Comprehensive Foundation", Macmillan College Publishing Company, N.Y., 1994.

6. K. Y. Kung, "Digital Neural Networks", PTR Prentice Hall, NJ, 1993.

7. M. Moller, "A Scaled Conjugate Gradient Algorithm for Fast Supervised Learning", Neural Networks, Vol. 5, pp. 525-533, 1993.

8. SNNS (Stuttgart Neural Network Simulator) User Manual, Version 4.1, 1995.

9. J. A. Kong, "Electromagnetic Wave Theory", 2nd edition, John Wiley and Sons, 1990.

10. H. Szu, B. Telfer, and S. Kadambe, "Adaptive Wavelets for Signal Representation and Classification", Opt. Eng., Vol. 31, pp. 1243-1256, 1992.

11. H. Szu, B. Telfer, and J. Garcia, "Wavelet Transforms and Neural Networks for Compression and Recognition", Neural Networks, Vol. 9, pp.695-708, 1996.

12. G. Kaiser, "A Friendly Guide to Wavelets", Birkhauser, 1994.

13. J. Stoer, R. Bulirsch, "Introduction to Numerical Analysis", 2nd edition, Springer-Verlag, 1993.

14. A.N. Kolmogorov, "Interpolation and Extrapolation of Stationary Random Sequences", Translated by the Rand Corporation, Santa Monica, CA, 1962.

15. Allen L. Edwards, "an Introduction to Linear Regression and Correlation", W.H. Freeman and Company, N. Y., 1984, pp. 98-99.

16. N.N Evtikhiev, B.N Onykii, D.V. Repin, I.B. Scherbakov, R.S. Starikov, M.I. Zabulonov, "Optical Hardware Implementation of the Two-layer Neural Network with the Pre-processing Unit for Invariant Pattern Recognition", SPIE Vol. 2752, pp. 281-289, Orlando, FL, 1996.

17. G. A. Korn, T.M Korn, "Mathematical Handbook", McGraw-Hill Book Company, N.Y., 1968.

18. Gail A. Carpenter, Lecture notes and handouts of the course CN 550, Cognitive and Neural Systems Department, Boston University, Boston, MA, 1997.

19. Joseph Plunkett, ECE Dept., WPI, personal communications, 1997.

20. URL: http://wwwipd.ira.uka.de/~prechelt/FAQ/neural-net-faq.html, the compilation of the materials of the Usenet newsgroup comp.ai.neural-nets (this site is mirrored on many different hosts, so it probably is not going to disappear from the WEB).

21. Rumelhart, D. E. and McClelland, J. L., Parallel Distributed Processing: Explorations in the Microstructure of Cognition, vol. 1, pp. 318-362 The MIT Press, 1986.

22. Sergey Perepelitsa, ECE Dept., WPI, personal communications, 1997.

23. Pavel Bretchko, ECE Dept., WPI, personal communications, 1997.

Appendix A SIMULATION RESULTS FOR SYNTHETIC GPR LINES

A.1 Training set

A synthetic training set (main training set) consisted of approximately 3300 artificial scans representing seven possible subsurface patterns was generated. Depths of the layers were varied within the limits exceeding possible depth variations for the synthetic lines based on the real geometry analyzed with GPR technique. The subsets of the training set corresponded to patterns 1, 2, 3, 5, 6, and 7 were used for training of the NN2 models for each of those patterns, number of examples in each subset varied from 300 to 600.

A.2 Synthetic GPR line based on CR93-11

The approximate state coordinates of the starting and the end points of the line are {(243137.0, 3968939.0),(247863.0, 3970911.0)}. Line direction is from west to east. The first 600 meters of the line were approximated with the synthetic geometry. Then artificial A-scans were decomposed into the set of Adaptive Transform shift and weight coefficients. The geometry, the dielectric permittivity values, and the subsurface patterns distribution (red capital "P" with numbers) are shown in Figure A-1.

Neural Network 1 trained with the main training set was used to identify subsurface patterns and corresponding layer depths for CR93-11 based synthetic line. Pattern identification results are sown in Figure A-2, depth predictions for patterns 1, 2 and 3 (using corresponding NN2 modules) - in Figures A-3, A4, and A-5.

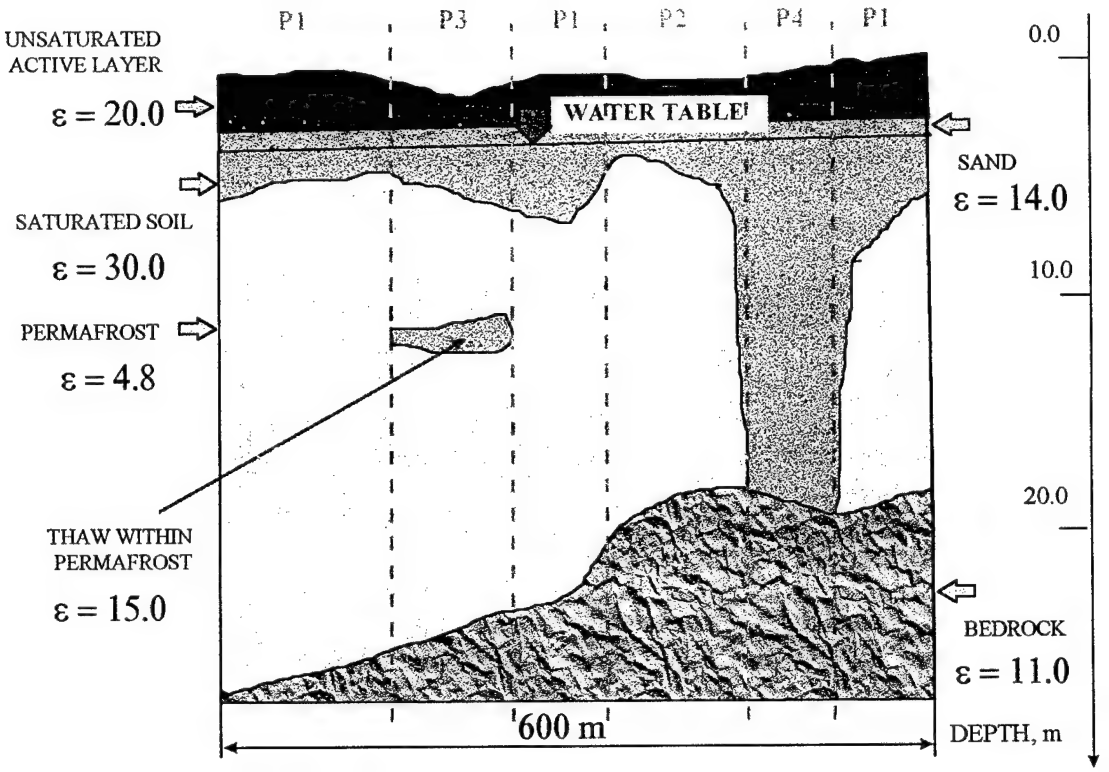


Figure Appendix A -1. Geometry and dielectric constants for synthetic CR93-11 GPR line.

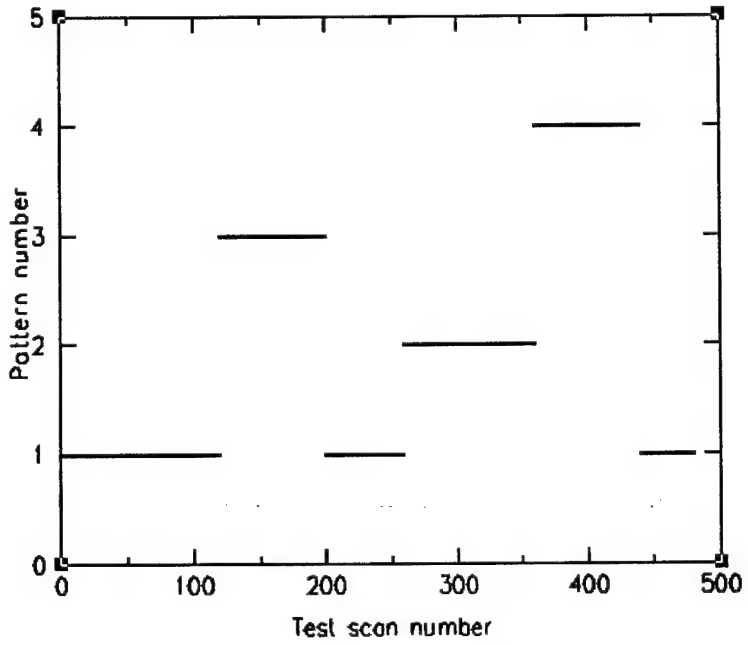


Figure Appendix A -2. Pattern identification results for CR93-11 GPR line.

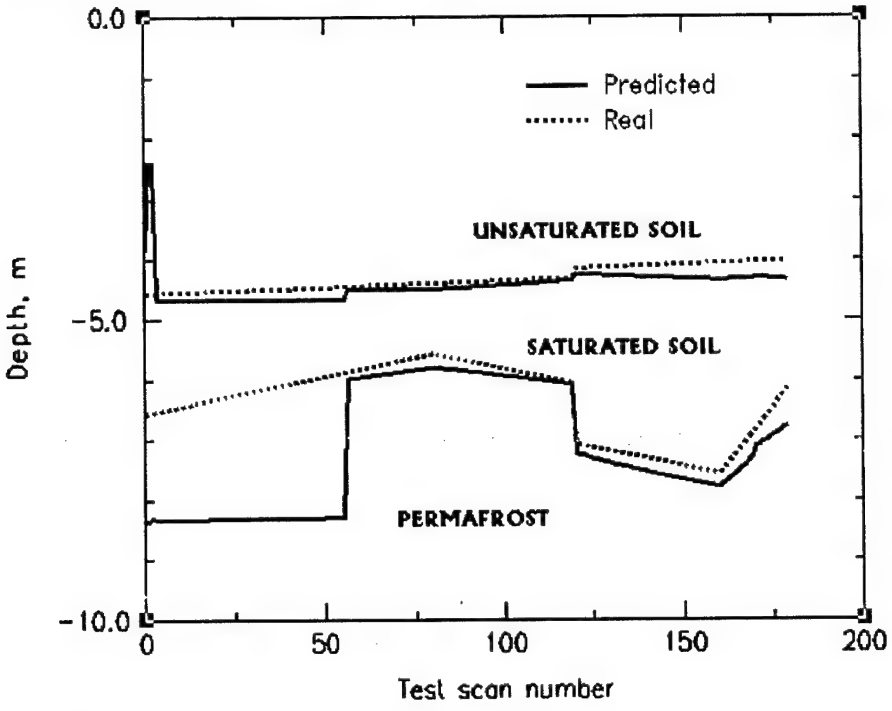


Figure Appendix A -3. Depth predictions for Pattern 1.

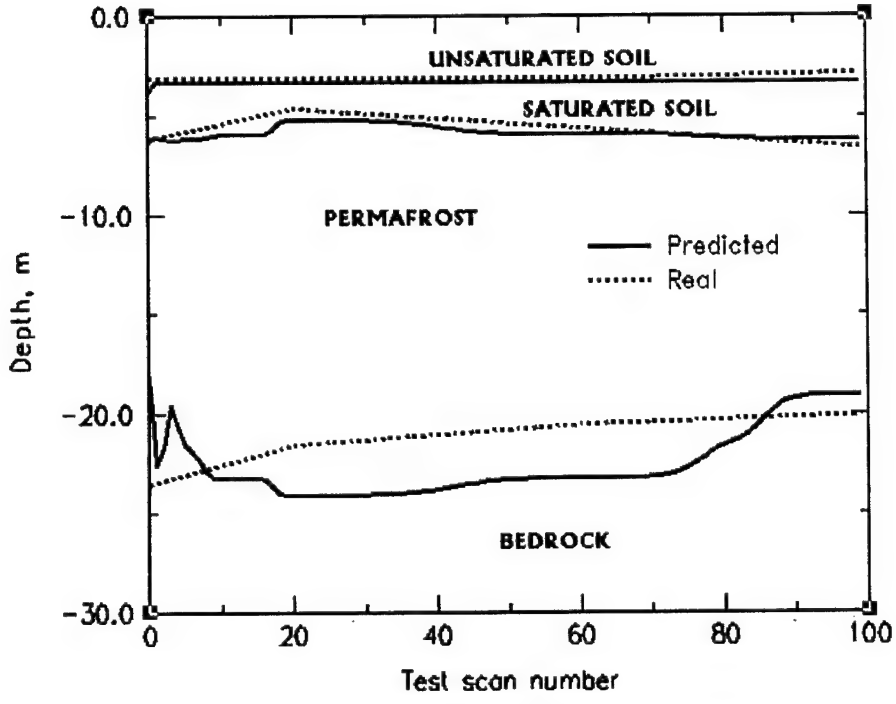


Figure Appendix A -4. Depth predictions for Pattern 2.

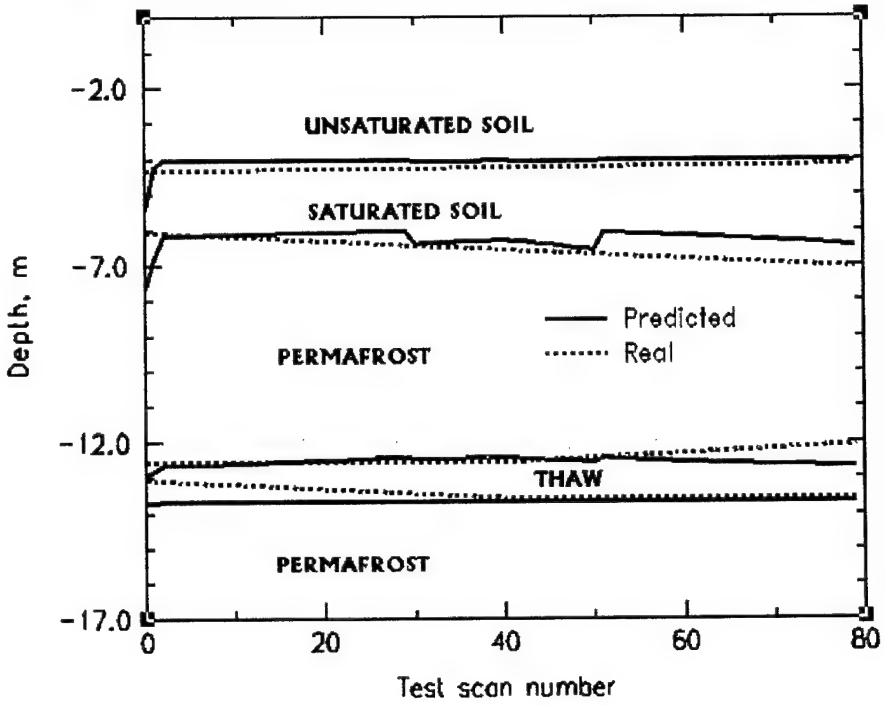


Figure Appendix A -5. Depth predictions for Pattern 3.

A.3 Synthetic GPR line based on CR94-61r

The approximate state coordinates of the starting and the end points of the line are {(246291.0, 3968477.0),(246266.0, 3970243.0)}. Line direction is from south to north. The geometry and the dielectric permittivity values are shown in Figure A-6, first 900 meters of the line were used.

Neural Network 1 trained with the main training set was used to identify subsurface patterns and corresponding layer depths for CR94-61r based synthetic line. Pattern identification results are sown in Figure A-7, depth predictions for patterns 5, 6 and 7 (using corresponding NN2 modules) - in Figures A-8, A9, and A-10.

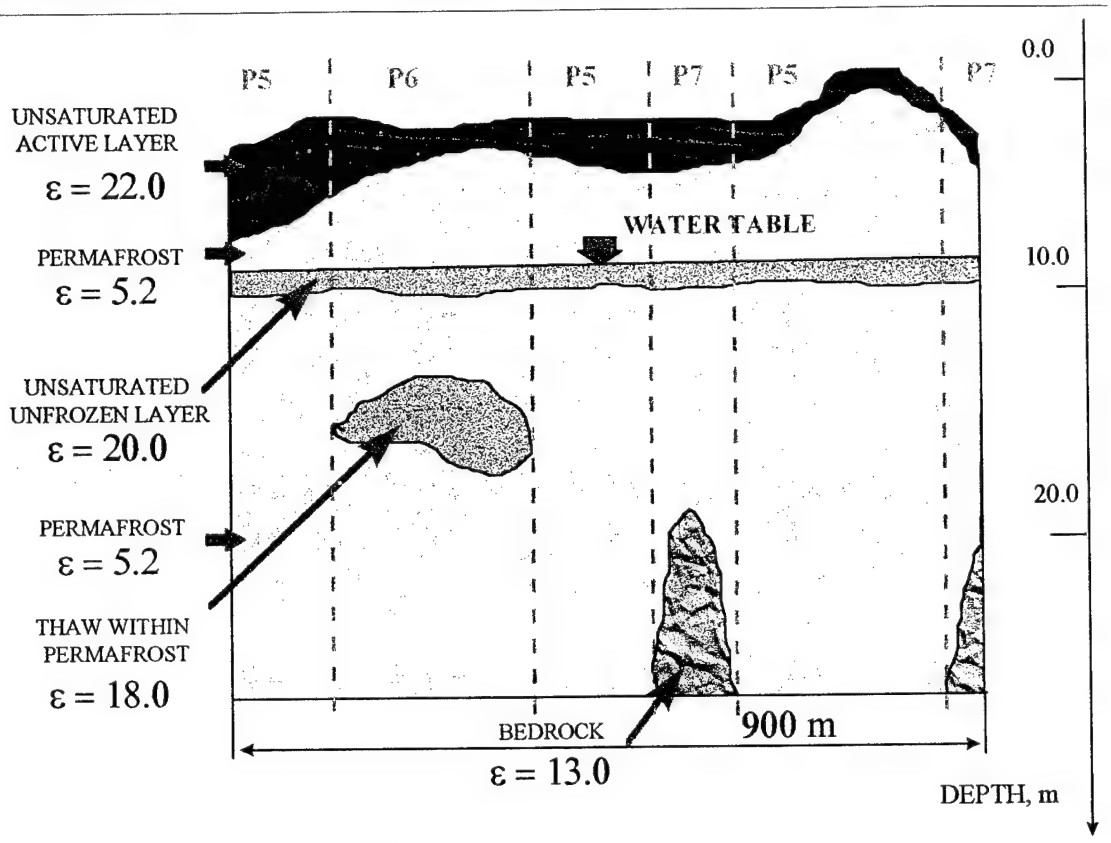


Figure Appendix A -6. Geometry and dielectric constants for synthetic CR94-61r GPR line.

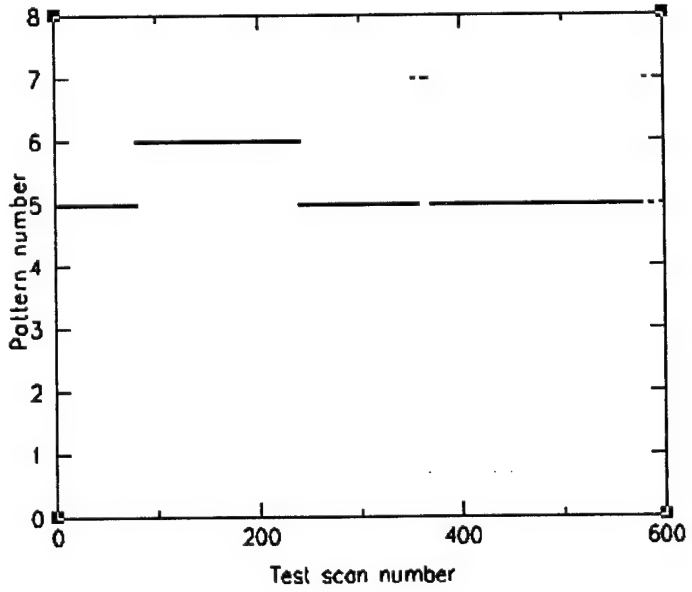


Figure Appendix A -7. Pattern identification results for CR94-61r GPR line.

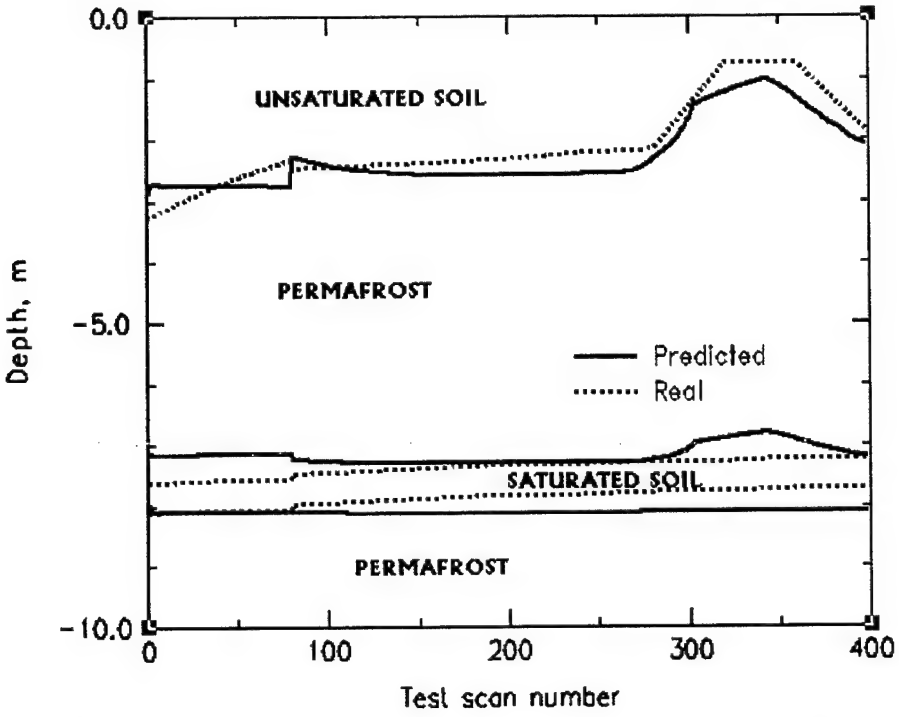


Figure Appendix A -8. Depth predictions for Pattern 5.

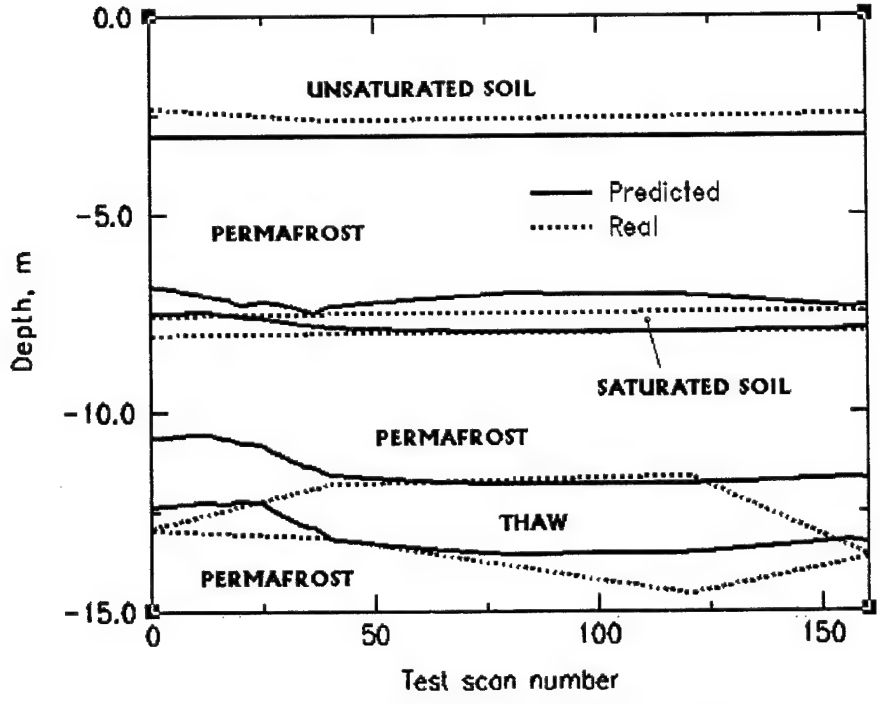


Figure Appendix A -9. Depth predictions for Pattern 6.

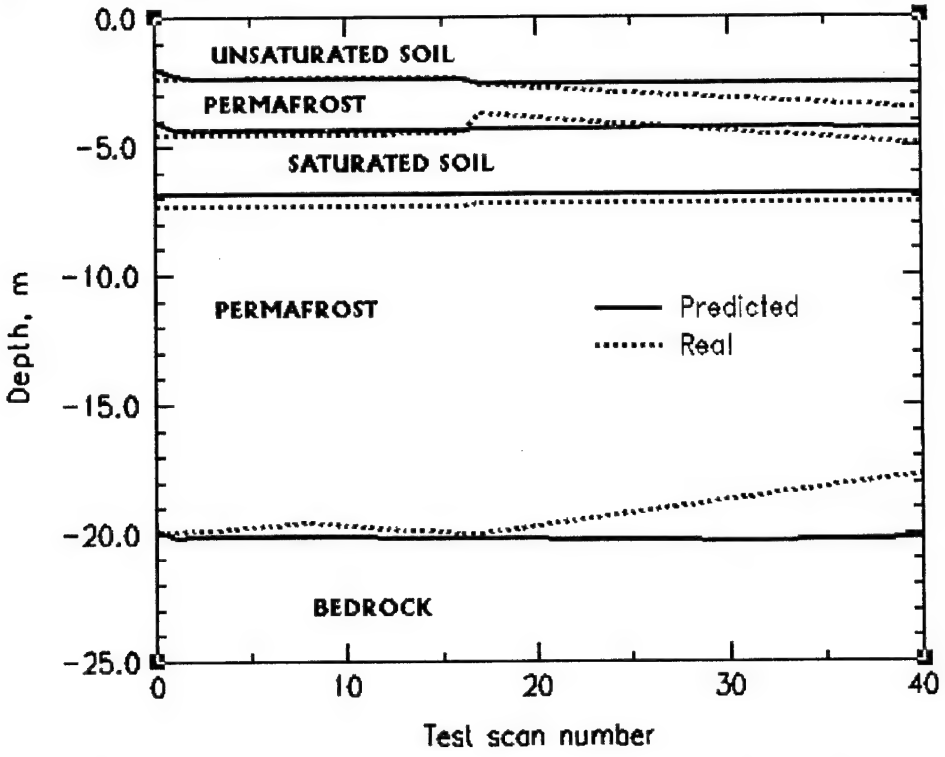


Figure Appendix A -10. Depth predictions for Pattern 7.

Appendix B. SUBSURFACE PATTERNS CLASSIFICATION

B.1 Pattern 1

Table Appendix B. .1. Layers configuration for subsurface pattern 1.

Pattern Number	Season	Layers	Dielectric constant
1	Autumn	Active layer (organics)	14.0 - 25.0
		Silt or sand (dry)	< 20.0
		Sand or gravels (saturated)	20.0 - 45.0
		Permafrost	4.4 - 5.6

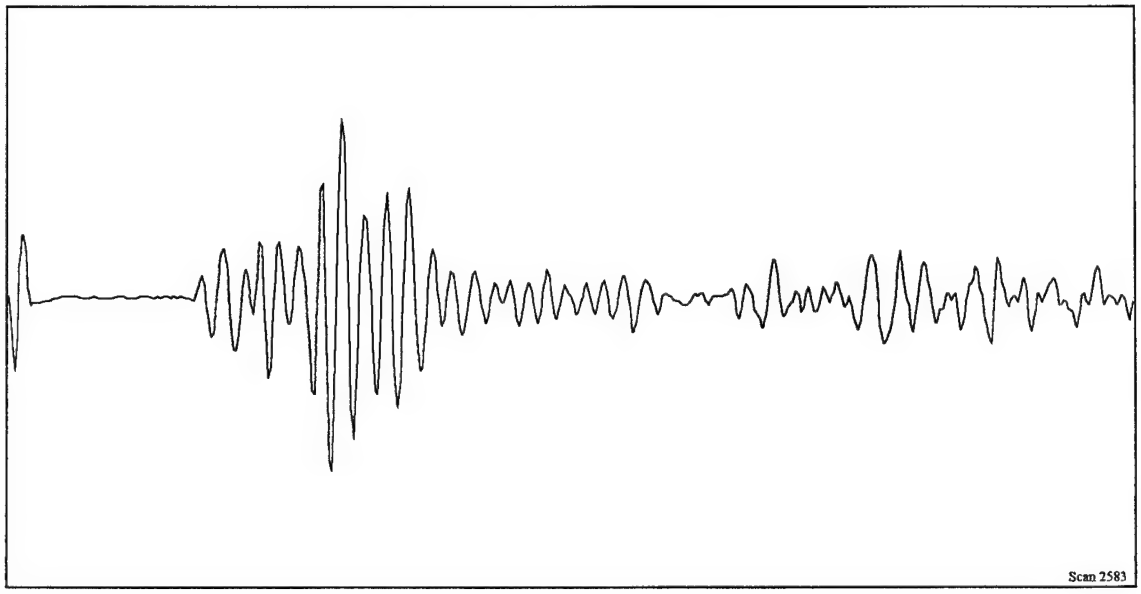


Figure Appendix B. -1. Typical A-scan from GPR line CR93-11 of subsurface pattern 1.

B.2 Pattern 2

Table Appendix B. .2. Layers configuration for subsurface pattern 2.

Pattern Number	Season	Layers	Dielectric constant
2	Autumn	Active layer (organics)	14 - 25
		Silt or sand (dry)	< 20.0
		Sand or gravels (saturated)	20.0 - 45.0
		Permafrost	4.4 - 5.6
		Weathered Bedrock	> 11.0

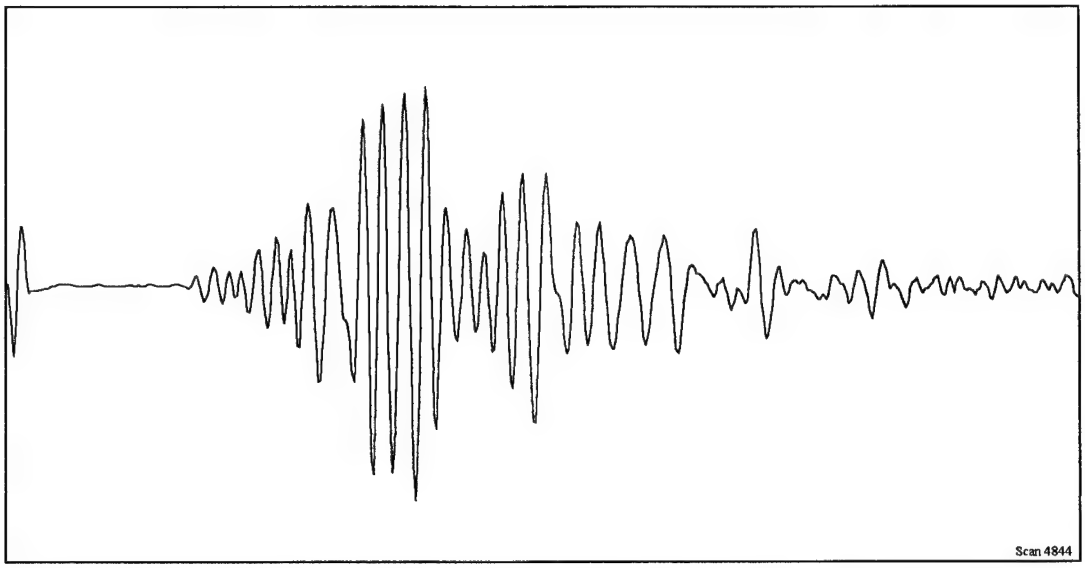


Figure Appendix B. -2. Typical A-scan from GPR line CR93-11 of subsurface pattern 2.

B.3 Pattern 3

Table Appendix B. .3. Layers configuration for subsurface pattern 3.

Pattern Number	Season	Layers	Dielectric constant
3	Autumn	Active layer (organics)	14.0 - 25.0
		Silt or sand (dry)	< 20.0
		Sand or gravels (saturated)	20.0 - 45.0
		Permafrost	4.4 - 5.6
		Cryopeg or Intrapermafrost diffractions (fine-grained sediments)	> 10.0

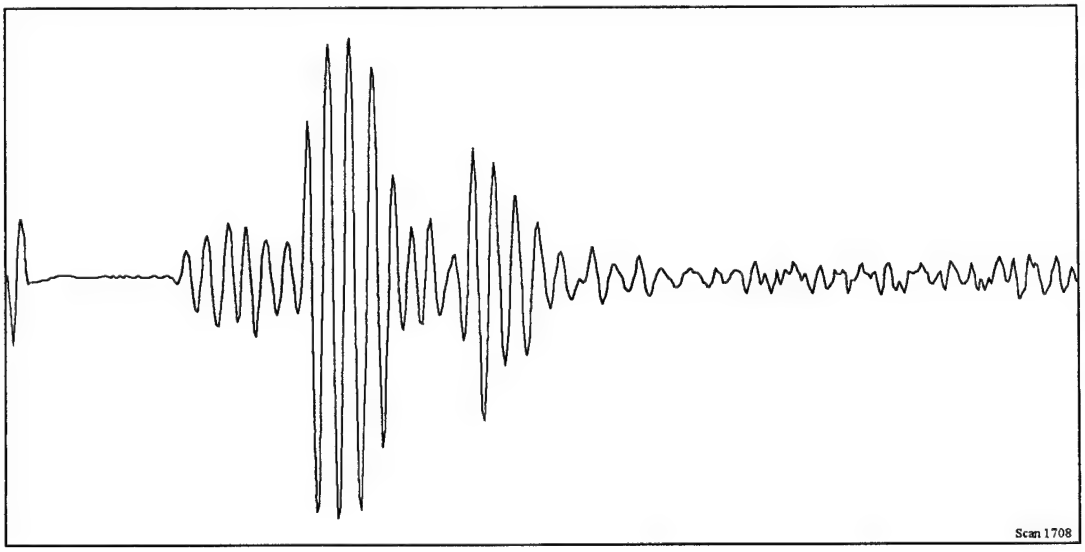


Figure Appendix B. -3. Typical A-scan from GPR line CR93-11 of subsurface pattern 3.

B.4 Pattern 4

Table Appendix B. .4. Layers configuration for subsurface pattern 4.

Pattern Number	Season	Layers	Dielectric constant
4	Autumn	Active layer (organics)	14.0 - 25.0
		Silt or sand (dry)	< 20.0
		Sand or gravels (saturated)	20.0 - 45.0
		Unknown layers	Unknown properties

No typical A-scan is available, the pattern was used to avoid the possible network confusion. Simulated scans in this area do not match the experimental GPR data, the main reason for introducing this pattern is to somehow label the data, that do not belong to any category currently used.

B.5 Pattern 5

Table Appendix B. .5. Layers configuration for subsurface pattern 5.

Pattern Number	Season	Layers	Dielectric constant
5	Spring	Active layer (organics)	14 - 25
		Permafrost	4.4 - 5.6
		Unfrozen sand or gravels (saturated)	20 - 45
		Permafrost	4.4 - 5.6

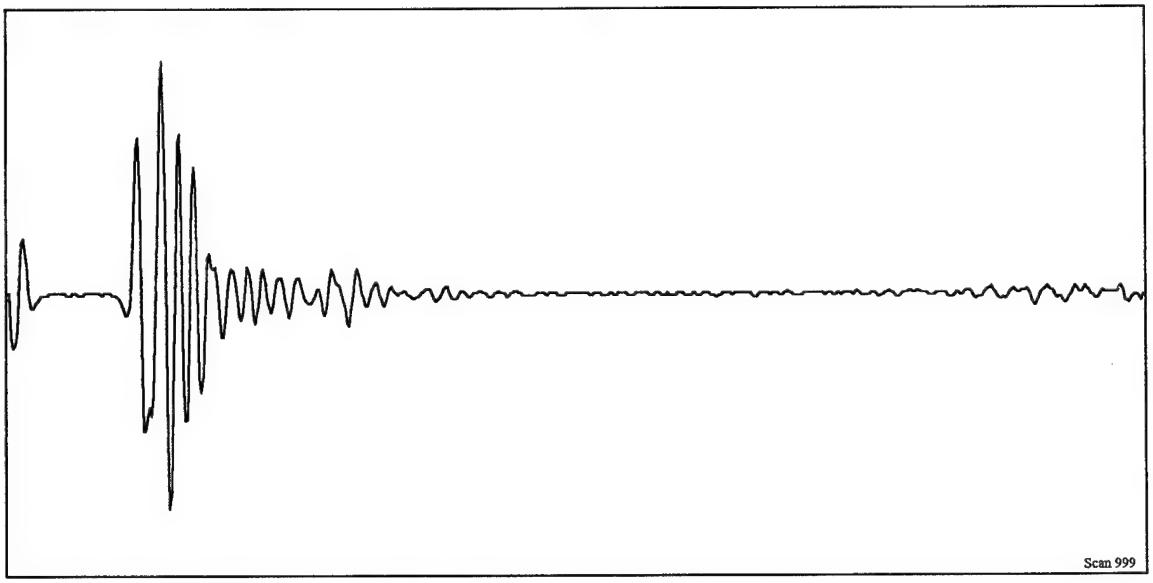


Figure Appendix B. -4. Typical A-scan from GPR line CR94-61r of subsurface pattern 5.

B.6 Pattern 6

Table Appendix B. .6. Layers configuration for subsurface pattern 6.

Pattern Number	Season	Layers	Dielectric constant
6	Spring	Active layer (organics)	14 - 25
		Permafrost	4.4 - 5.6
		Unfrozen sand or gravels (saturated)	20 - 45
		Permafrost	4.4 - 5.6
		Cryopeg or Intrapermafrost diffractions (fine-grained sediments)	> 10.0

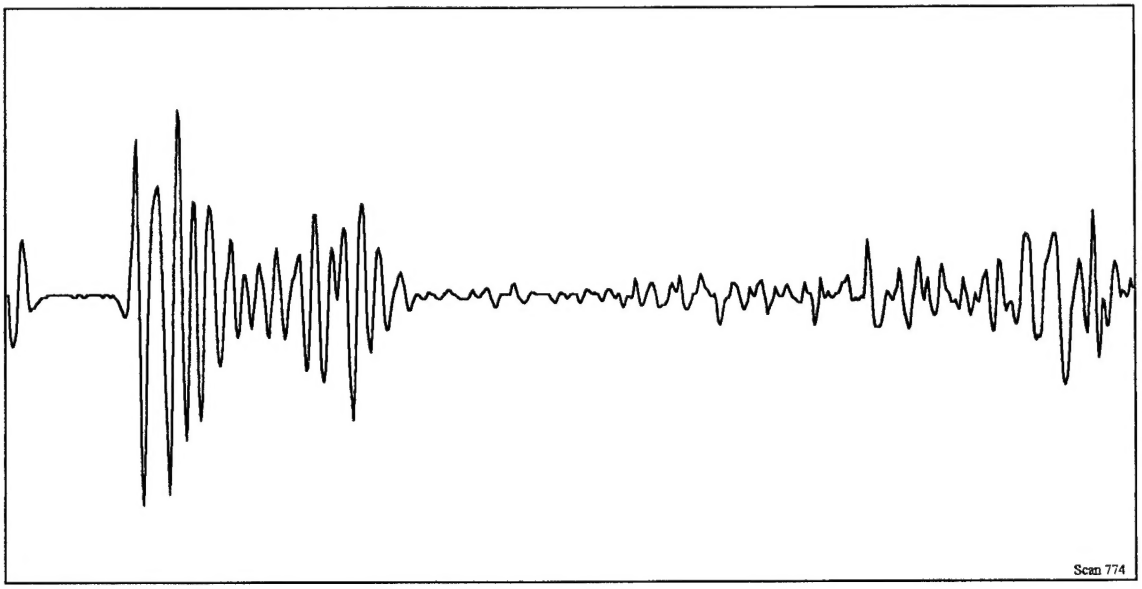


Figure Appendix B. -5. Typical A-scan from GPR line CR94-61r of subsurface pattern 6.

B.7 Pattern 7

Table Appendix B. .7. Layers configuration for subsurface pattern 7.

Pattern Number	Season	Layers	Dielectric constant
7	Spring	Active layer (organics)	14 - 25
		Permafrost	4.4 - 5.6
		Unfrozen sand or gravels (saturated)	20 - 45
		Permafrost	4.4 - 5.6
		Weathered bedrock	> 11.0

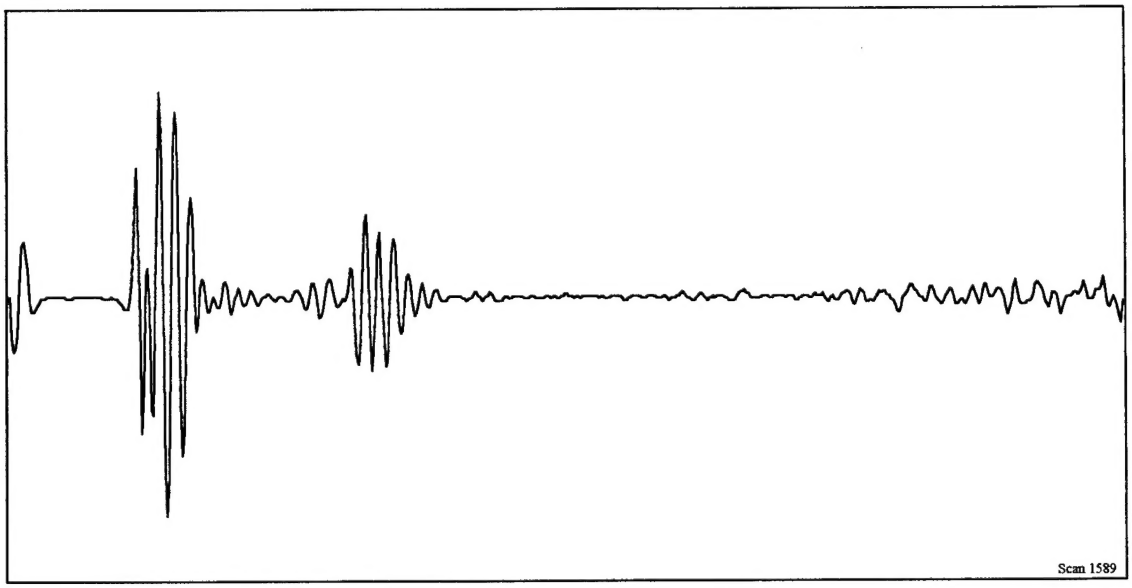


Figure Appendix B. -6. Typical A-scan from GPR line CR94-61r of subsurface pattern 7.

Appendix C SIMULATION RESULTS FOR REAL GPR LINE

Neural Network 1 trained with the main training set (Appendix A) and six Neural Network 2 modules corresponding to patterns 1, 2, 3, 5, 6, 7 were used to estimate the patterns and depths for the real GPR line CR93-11. The results are shown in Figure C-1. Numbers above y = 0 denote the subsurface pattern number, below y = 0 - correspond to the interface depth values.

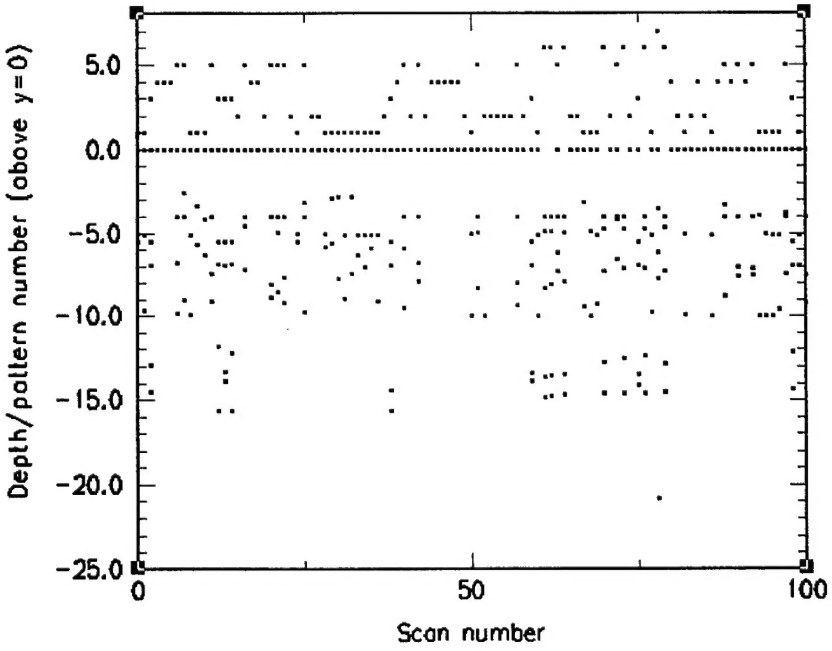


Figure Appendix C -1. Pattern and depths identification results for the real GPR radar line CR93-11.

The data in the above figure are somewhat confusing due to the pattern change observed in the upper part of the figure. If one attempts to connect several of the points to sketch out the layer depths, the following picture would be produced - see Figure C-2.

Note, that not all the points were connected due to uncertainty of the results as explained in the main text of this report.

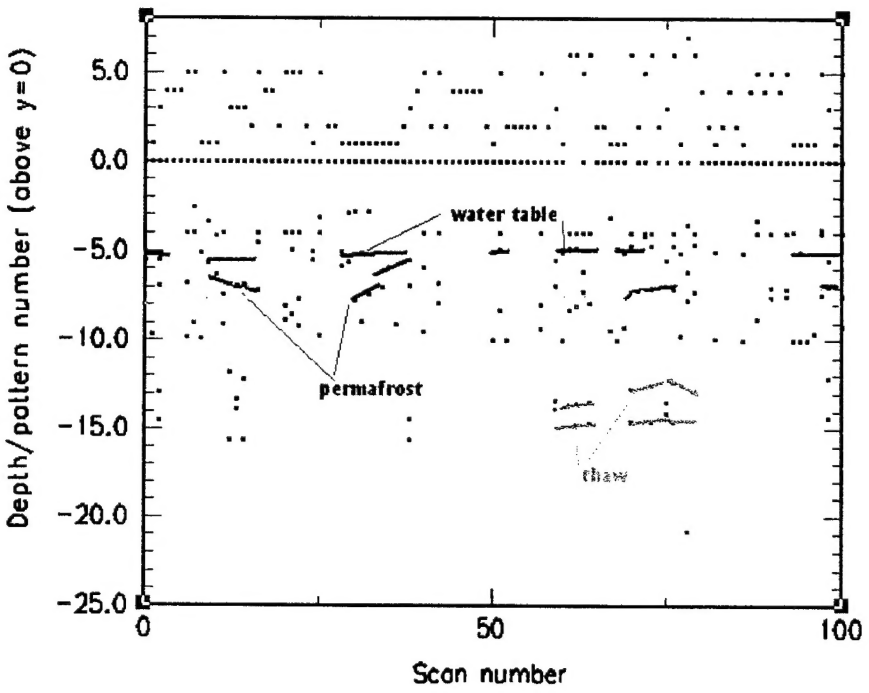


Figure Appendix C -2. An attempt to connect the adjacent points to produce layer interfaces.